

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2002, 23:25:50 ; Search time 2520.32 Seconds
(without alignments)
5965.754 Million cell updates/sec

Title: US-09-728-309-1
Perfect score: 1114
Sequence: 1 attcagcatcatagatgt.....gagagggttaaccattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_estli:*

10: gb_est2:*

11: gb_hic:*

12: gb_gss:*

13: em_gss_hum:*

14: em_gss_inv:*

15: em_gss_pln:*

16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	199.6	17.9	725	12	AQ078448	CIT-HSP-2
2	188.2	16.9	690	12	A2840584	2M0138N11
3	184.6	16.6	777	12	A2741291	RPCI-24-9
4	159.6	14.3	808	12	BH039878	BH039878 RPCI-24-2
5	153.2	13.8	729	12	BH083291	RPCI-24-3
6	141.8	12.7	617	12	A2795382	2M0049N06
7	141	12.7	545	12	A2748309	RPCI-24-6
8	135.2	12.1	635	12	BH062438	BH062438 RPCI-24-3
9	132.2	11.9	574	9	A1806860	wf36b07.x
10	129.4	11.6	615	12	A2553624	RPCI-23-2
11	129.4	11.6	610	12	A2222872	RPCI-23-4
12	128.4	11.5	467	12	AQ117475	HS_2187.B
13	127.4	11.4	563	12	A2810472	2M0076E06
14	127	11.4	638	12	A2553591	RPCI-23-2
15	126.8	11.4	740	12	BH267496	CH230-82L
16	119	10.7	853	12	A2265296	RPCI-23-1
17	118.2	10.6	493	12	A2759046	1M0551P17

ALIGNMENTS

18	116.4	10.4	447	9	AA012849	
19	116.4	10.4	768	12	BH060668	
20	114.2	10.3	597	12	A2422002	RPCI-24-3
21	113.4	10.2	715	12	AG013852	Homo sapi
22	112	10.1	643	12	BH048022	AG013852 RPCI-24-2
23	112	10.1	698	12	BH039875	RPCI-24-2
24	111.8	10.0	397	9	AA021085	AA021085 z665408.r
25	111.2	10.0	424	10	H86939	ys72f11.r1
26	110.2	9.9	586	12	A2093307	RPCI-23-4
27	108.6	9.7	717	12	AG013832	Homo sapi
28	108	9.7	682	12	AG130179	Pan trogl
29	105	9.4	511	12	B67083	CIT-HSP-201
30	104	9.3	495	12	A0223006	HS_2017.A
31	100	9.0	243	12	A2737781	RPCI-24-8
32	98.4	8.8	716	12	AG013834	Homo sapi
33	97.8	8.7	715	12	A2994117	2M0279M22
34	96.4	8.7	696	10	BE968588	601649752
35	95.6	8.6	725	12	AG013831	Homo sapi
36	94.6	8.5	382	9	AA442630	zV59f10.r
37	92.8	8.3	677	12	BH267500	CH230-82L
38	91	8.2	580	9	AW973537	EST385637
39	90.4	8.1	481	12	B92737	CIT-HSP-217
40	87.8	7.9	406	12	A2767149	1M0566H11
41	86.8	7.8	597	12	A2896363	RPCI-24-2
42	86.4	7.8	593	12	A2754210	RPCI-24-6
43	84.6	7.6	207	12	B59457	CIT-HSP-201
44	84	7.5	658	12	AQ666848	HS_2105.A
45	83	7.5	675	12	AG000745	Homo sapi

/note="vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"
BASE COUNT 190 a 184 c 120 g 231 t

Query Match 17.9%; Score 199.6; DB 12; Length 725;
Best Local Similarity 63.0%; Pred. No. 8.4e-38;
Matches 345; Conservative 0; Mismatches 194; Indels 9; Gaps 2;

QY 15 agatgttgaaattggtattattgagacaatgagcagaaattatgctattctcattagatc 74
Db 76 AAAAGATGATGATGGTAAACAAATTTGAACACATGAACCAACCAACATTAACATGGGATC 135
QY 75 tcttgctt-----ttctcacagatatccttctgctttaaatttcttcccttcaagatgaca 128
Db 136 TATTGTTTCTTTTATCCAGATATAGAAAGATGTTTCTCTATGCCATTAAAGATGAAC 195
QY 129 aactctcgtgtttattacatacaaatctcttctttttccacagccagcttttggaatt 188
Db 196 TATTTCTACTTATGCTATCGTTAAAAATGCCTTTGTTTTCRAAGCTGGCATTTGGATTCT 255
QY 189 cagcaaacacacacctctctctttccacatcttcacctttgtttttcagtcacaggtcta 248
Db 256 CAGGCACACCTTTCTCTCTTTTCTCCACATCTTTAGCTCTCTCTGAATCGCAGGCCTA 315
QY 249 agtcattgacatgataattagtagcctgtctctctcaccacatactgctctctcactc 308
Db 316 AACCCGTGACTTACTCACCTGTCACCTGGCCCTCAFTTCACATTCAGATGCTCCTCACTG 375
QY 309 agcaaatatgggtccttagactcttttgggttcacagaatactcagatgatcttaggt 368
Db 376 CAGTGGATTTTGGCCCTAGACATATTTGAATCACTGCAATTTTGGGAATGACTTTAAGT 435
QY 369 ataaggctcatgtctttttaaacaagtgatgagggccctctccatctgcacccctgccc 428
Db 436 GTAAGGCACATTTTACACAAACAGGGCAATGAAGGCCCTCTCACTGTCACCACTGCC 495
QY 429 tctcgtgctgctccagc---catcatcagcccccagcatctctctcttgccgaagctca 485
Db 496 TTCTGAACATGCTCCAGGCCATCATCATCAGCCCCAGCACTTCTGGTTGGCAAGGTTA 555
QY 486 acactctctcgaagtcacatcttaggattctctcttctctcatggtgctcacaatgt 545
Db 556 AACATAATCCAAATATACATTTTGCATGTTTCTCTTTGGGTGCTGCTCAATTTGT 615
QY 546 tcattggt 553
Db 616 CTTCACT 623

RESULT 2
AZ840584
LOCUS 2M0138N11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M0138N11 F, DNA sequence.
ACCESSION AZ840584
VERSION AZ840584.1 GI:13010492
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 690)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mus whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0138 row: N column: 11
Seq primer: CGTGTAAACAGCAGCGCCAGT
Class: plasmid ends
High quality sequence stop: 690.

FEATURES
source

1. .690
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0138N11"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: pWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (GI:4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 174 a 184 c 118 g 214 t
ORIGIN

Query Match 16.9%; Score 188.2; DB 12; Length 690;
Best Local Similarity 60.6%; Pred. No. 4.8e-35;
Matches 396; Conservative 0; Mismatches 248; Indels 9; Gaps 5;
QY 149 atacaatctcttttatcacaaagccagcttgggaatttcagcaaacacacatctctt 208
Db 33 ATAGGAAACAAATTTTCTCTGAAATGGCATTTGGGATCTCAGGCAACAGCTTCCCTCTT 92
QY 209 cttttccacatcttcaccttgttttttcagtcacaggtcttaagtcattgacatgataatt 268
Db 93 CTGTGTCACATCTCTAAAGTTTCATTCGTGGCAGACAGCCAGACTCACTGACCTGCCATT 152
QY 269 agtcacctgtctctcatccacatactgctctcttctcactcaggcaatatattggtcctta 328
Db 153 GGTCTCTTGTCCCTAAATCCACCTACTGCTACT---GGTCGACGATTCATAGGCACA 209
QY 329 gactcttgggttcacagaatactcaggtatgcttaggtatataaggtcattgctctttta 388
Db 210 GACATTTTATCTCTCGAGGGGCTGGGATGACATCATATGTAAATTCCTGTGTGCTCTG 269
QY 389 acaaggtgatgaggggctctccatctgcacccctgcctcctgctgagtggtccaggcc 448
Db 270 TACAGAGTTTGAAGGGTTCTCCCTTTCACCCAGCAGCATGTTGAGTATCTCTCAGGCC 329
QY 449 atca-----tcagccccagcatcttctctcttgccaaagctcaaacatctctctgcaagtcac 505
Db 330 ATCATTTCTTAGTCCCAAGAGCTCTCTGTTTAGCAAAAGTTCAAGCATATATCTCTCATCAC 389
QY 506 atcttaggattcttctcttctctcatgggtctcacaatgcttgggt-gtaattctctg 564

Db	390	ATATCAAGTGCCATCTTTTCTGAGTGTCTCTATATGTTAATTGGCAGTCAACTCTTG	449
QY	565	ctgtacactgagggtacacccacggaacggggccagctctctgtt-tgtcatcacagcac	623
Db	450	GTATCCATCATTTGCCACCCCAATTTGACCATGAATGACTTTATTTATGTTACTCAGTCC	509
QY	624	tgttctcttttggccatgagctacacccacagaggagactgttttttcacacactaataagctttg	683
Db	510	TGCTCTATTCTACCCCTTGAGTGTACTCATGCAAGAGCATATATTCTACACTGCTAGCCCATC	569
QY	684	aggatgtcacacttttagggttcatggtctctcagagagctacatgggtgattattttta	743
Db	570	AAGCAATCTTTCTATTAGTCTCATGCTCTCTCTAATTGGTACATGGTAGCCCTCTTA	629
QY	744	tacagacaataaagggttatctcagttcagttcagtcagcagccagctgtcccccga	796
Db	630	AGCATGC-ACAGGAACAGACACAGCATCTTTCATGGAACCAACCTGTCCCCCA	681
RESULT 3			
A2741291/c			
LOCUS	A2741291	777 bp	DNA linear GSS 25-JAN-2001
DEFINITION	RPCI-24-90L5.TV RPCI-24 Mus musculus genomic clone RPCI-24-90L5, DNA sequence.		
ACCESSION	A2741291		
VERSION	A2741291.1 GI:12516500		
KEYWORDS	GSS.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 777)		
AUTHORS	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.		
TITLE	Mouse BAC End Sequences from Library RPCI-24		
JOURNAL	Unpublished (1999)		
COMMENT	Other_GSSs: RPCI-24-90L5.TV Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html Plate: 90 row: L column: 5 Seq primer: SP6 Class: BAC ends.		
FEATURES			
source	1. .777 /location="Qualifiers /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="RPCI-24-90L5" /clone_lib="RPCI-24" /sex="Male" /cell_type="Spleen/Brain" /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."		
BASE COUNT	260 a	141 c	154 g 222 t
ORIGIN			
Query Match	16.6%; Score 184.6; DB 12; Length 777;		
Best Local Similarity	60.9%; Pred. No. 3.7e-34;		

Matches	388;	Conservative	0;	Mismatches	239;	Indels	10;	Gaps	5;
QY	385	tttaaacaaagtgatgagggccctccatctgcacccccctgcctcctgagtggtccca	444						
DB	774	TTTAAACAGGGCAATGAGAGGTCTCTATCAGCATCACCTGCCTCTAGAGTGTGCCAA	715						
QY	445	ggccatc---atcagccccagcatctctccttgggcaaaagctcaaacatcctctctgaag	501						
DB	714	GGCTGTGACGATCAGTCCCAACTCCGCTTTGTTGGCACAAATTTAAACATAAACAATAAAAA	655						
QY	502	tcacattctaggatctctcttcttcattgggtccctcaacatgttcatctggtgtaactctt	561						
DB	654	ATACATGATCTATTCTTCTCTATTGCTCTTCAATTTGCTCTTCAATTTGCTCTATGACATTAACAG	595						
QY	562	ctgct--gtacactgcggctaccctccagtgaaacggggccagttctcttctgttgcataca	619						
DB	594	GATCTTCTATACTGGTGTATATCAACGCTGAGTGAGACCAACGAGTGAAGGTCACATAA	535						
QY	620	gcactgttctcttttgcctatgagctacacccacagagagactgtttttcacactaatgac	679						
DB	534	ATACTGCTACTCTTCCCTATGAACAACATCATCAGGGGACTAATTTCAACAATCATTAAC	475						
QY	680	tttgagggatgtcacctttatagggttcatgtctctcctcaagagggctacatggtgattat	739						
DB	474	CTTAAGAGATGTATTTTGTAGGAGTCATGCTGAGCACAAGTACATACATGGTGATTAT	415						
QY	740	tttatcacagaataaagggtctatctcagtccttcacgcagccagcctgtccccagagtc	799						
DB	414	CTTGTTCAGAC-ATCAGAGGCAATGCAAGCATCTTCATAGCGACAGCCACTT-GAGAGCC	357						
QY	800	tcaccagtgaagaagcctccagggctatcttactgtgtgagttt---tgtcttcaca	856						
DB	356	TCCCTTGAGAAAAGAGCCACCAGACCATCTTGCTGCTGGTGTCTTCTATGTGGTCAAG	297						
QY	857	tactgggtggactttaagttctcttattttcaggaggtgtgacatggataaagattctctg	916						
DB	296	TACTGGGTGGACTTTATATGTCATCTAGAACAGTCTTTTATGATGATGATGACCCAGTC	237						
QY	917	ctagtgtggtccagggtattgttgcccaatagctatgcgcgaatagtcctcttgatgcta	976						
DB	236	ATGCTTACTGTTCAGAAAGTTTGTGTTGAATGCCCTATCCCAACAATTACTCTCTTAATACAA	177						
QY	977	attatgtctgataccaaatattcaagactctgcaaa	1013						
DB	176	ATTAGTCTGTATGAATCGAATACTCATGATTCTTCAAAAA	140						
RESULT 4									
BH039878									
LOCUS	RPCI-24-273F22.TV RPCI-24 Mus musculus genomic clone RPCI-24-273F22								
DEFINITION	, DNA sequence.								
ACCESSION	BH039878								
VERSION	BH039878.1 GI:14818265								
KEYWORDS	GSS.								
SOURCE	house mouse.								
ORGANISM	Mus musculus								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 808)								
AUTHORS	Zhao,S., Nierman,W., Malek,J., Shvartsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.								
TITLE	Mouse BAC End Sequences from Library RPCI-24								
JOURNAL	Unpublished (1999)								
COMMENT	Other_GSSs: RPCI-24-273F22.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org								

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tadb/bac_ends/mouse/bac_end_intro.html. Plate: 273 row: F column: 22

Seq primer: T7

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..808

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-273F22"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 234 a 172 c 139 g 263 t

ORIGIN

Query Match

14.3%; Score 159.6; DB 12; Length 808;

Best Local Similarity 55.9%; Pred. No. 4.3e-28;

Matches 389; Conservative 0; Mismatches 294; Indels 13; Gaps 4;

QY 328 agactcttgggttcacagaataactcaagatgatctcttaggtataaggtcaattgtctctttt 387

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 39 AAACATATTAGAGTCACATAAAATTTGGGAATGGCATCAAAATGTAAACAACTTTTCCAT 98

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 388 aacaagatgatggggcctctccatctgcaccccctgctctctgagtggtctccaggc 447

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 99 AAACAGATGATGAGAGCCCTCTATCTGCATCACCTGCTCTTGAGTGATCCAGGC 158

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 448 cacc---atcagcccccgcattctctctctggaagcctcaaacatcctcttgcgaagtc 504

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 159 TATCCCAATCAGTCCCACTACCTCTTTGTCGCAAAATTTAAGAGATAAGTAATAAACA 218

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 505 catcttaggattctctctctctctctctcctcaacatcttctggtatctctctg 564

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 219 GATCATCTACGCTTTTCTATCTCTGATCTTTCATTTGT--CATTCAGTAGTAGCGG 276

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 565 ctgtacactcggcgtaccgccagtgaaacggggccagctctctctgtgtctacacagcact 624

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 277 ATCTTCCATGTTGCTGATTTTACCAATGTGAAACCTACCAGATGAAGTCACTAAATCCT 336

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 625 gttccttttgcccatgagctacacccacagggagactgtttttcacactaatgacttga 684

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 337 GCTCACCTTTCCCATGAACATACATCATCAGGGGGTGTGATTTAAACAGTAGCAATGTCCA 396

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 685 gggatgtcacctttataggttcatgtctctcctcaagaggctacatgtgattatttat 744

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 397 GAGATATATGCTTGTAGGAGTCACTGATGATCAAGATACATCATCATGCTGTTATCTTAT 456

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 745 acagacataaagggtctatctcagtcgtcttcacgacccagcctgtccccaggtctcacc 804

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 457 GCAACATCAGCAATGCAAGCAATTTTCATAGCATCACCA-----CCTGAGAGCATCTCC 511

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 805 agtgaagaagcctccaggctctcttactgctggtgagt---ttgtcttccacatactg 861

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 512 TGAGAAAAGGCCACCCAGACCATCTTGTGCTAGTAGTGTTCCTTTCTGTTTGTGTTACTG 571

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 862 ggtgactttacgttctcattttcaggagggtgtgacatggataaaatgattctctgtagt 921

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 572 GTTGGACTTCATCATCTCATCTACCTCAATCCTGTTTGTGATGTATGATCAGTCATCTCT 631

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 922 gtggctccaggttatgttgcccaatagctatgcccgaattagtccttctgctgtaattta 981

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 632 GACCATTACAGAGTTTGTGATGTATGCTATCCCAAAATTTGCTCCTTTGGTACAGTCAAG 691

QY 982 tgcgtataacacaaatatttcaagactctgcgaattgt 1017

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 692 TTCTGATAAGAGAATATCCGTATCTACTAAAAATTT 727

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5

BH083291

LOCUS

RPCI-24-309J17.TV RPCI-24 Mus musculus genomic clone RPCI-24-309J17

DEFINITION

BH083291

VERSION

BH083291.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 729)

AUTHORS

Zhao,S., Nierman,W., Malek,J., Shatsman,S., Aklnret,B., Levins,M.,

Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,

Russell,D., de Jong,P. and Fraser,C.M.

TITLE

Mouse BAC End Sequences from Library RPCI-24

JOURNAL

Unpublished (1999)

COMMENT

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end

page: http://www.tigr.org/tadb/bac_ends/mouse/bac_end_intro.html

Plate: 309 row: J column: 17

Seq primer: T7

Class: BAC ends.

FEATURES

Location/Qualifiers

1..729

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-309J17"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 170 a 181 c 138 g 240 t

ORIGIN

Query Match

13.8%; Score 153.2; DB 12; Length 729;

Best Local Similarity 58.2%; Pred. No. 1.5e-26;

Matches 384; Conservative 0; Mismatches 263; Indels 13; Gaps 6;

QY 78 tggcttttccacagatatctcttcttaatttcttcttaagatgatcaaaactctctg 137

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 74 TGTCTTTTCTCTCTGTTCTACATGTTCTCACATACATGAATGAACAGACGAGACTACATA 133

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 138 gttttattaccatacaaaattcttttaccacagccagcttttgaatttcagcaaaaca 197

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 134 CTAAATCTAAATCTGGGAAC-ACATTTTCTCTGATTTTCTTGGCATCACAGAAACA 192

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 198 ccacctctctctttccacatcttccaccttctgtttttcagtcacagctctcaagccattg 257

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 193 GCTTGCTTCTCTCTACACCATCTCAAGTTCATTCGTGGGCACAGGCCGAGACTCACTG 252

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DEFINITION RPCI-24-63G18.TV RPCI-24 Mus musculus genomic clone RPCI-24-63G18,
DNA sequence.
ACCESSION A2748309
VERSION A2748309.1 GI:12533043
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 545)
REFERENCE Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
TITLE Other_GSSs: RPCI-24-63G18.TJ
JOURNAL
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pjejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 63 row: G column: 18
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..545
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-63G18"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site:1: BamHI; Site:2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 160 a 116 c 112 g 157 t

Query Match 12.7%; Score 141; DB 12; Length 545;
Best Local Similarity 60.9%; Pred. No. 1.2e-23;
Matches 333; Conservative 0; Mismatches 205; Indels 9; Gaps 6;
QY 301 cttcaactcagcaaatattggtccttagactcttttggttcacagaataactcaggatga 360
DB 1 CCTCATTGGAGGAATGTCATGCTGCAGACATATTGAGTCACATGAATGTCAGAAATGA 60
QY 361 tcttaggtataagatgcattgtctttttaacaaggtgatgagggccctccatctgcac 420
DB 61 CATCAATGTATAGCAACTTTATACACAACAGGGTGATGAGAGGCCCTCTCTATCAGCAT 120
QY 421 cccctgctcctgagtgtgtccaggc---catcatcagcccccagcatcttctccttggc 477
DB 121 CACCTGCCTCCTGAGTGTGATCCAGGCTGTGCACAAATCAGTCGCCGATACCTTTATGGTGCA 180
QY 478 aagactcaa-acatcctcttcgaagtcacatcttagattcttctcttcttcattggttc 536
DB 181 AAATTTAAACATAAATCAAGAAACACATGTCATGCTCTCTTTTATATTTGGTCT 240
QY 537 tcaecatgt--tcatgtggtgatctctgtgtacactgcgctacccccagtg-aaac 593
DB 241 TTCAATTTTCTCTTAGTAGTACCTGTGATCTTCTATACTGCTGGTTTACCAATGTGAGT 300

QY 594 ggggcccagttctctgttcttcacacagcactgttcttggccatgagctacacccac 653
DB 301 GAGACCAAGCAGATGAAGATCACTAAGTCTGCTCAATCTTGCCAAATGAACATACATCATC 360
QY 654 aggagactgttttcacacataatgactttgagggatgtccacctttatagggttcatggtc 713
DB 361 AGGGGAATGTTGTTAAACAGTGCACAACCGTCGAGAGATGTTTCTTGTGGGAGTTATCTG 420
QY 714 ctctcaagaggtcacatgttgattattttatcacagacaataagaggtctatctcagtgct 773
DB 421 ATCACAAGTCATACATGGTGATATCTTGTTCAGAC-ATCAGAGGCAATGCAAGCATCT 479
QY 774 tcacgcagccagcctgtcccccaggtctccacagtgaaagagcctccacgctatcttac 833
DB 480 TCATAGCATCA-ACAACCTGAGAACCTCCCTGAGAAAAGGCCACACAGACCATCTTGC 538
QY 834 tgctggt 840
DB 539 TGCTAGT 545

RESULT 8
LOCUS BH062438 635 bp DNA linear GSS 18-JUL-2001
DEFINITION RPCI-24-327B18.TJ RPCI-24 Mus musculus genomic clone RPCI-24-327B18
DNA sequence.
ACCESSION BH062438
VERSION BH062438.1 GI:14875948
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 635)
REFERENCE Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pjejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 327 row: B column: 18
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..635
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-327B18"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site:1: BamHI; Site:2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 174 a 130 c 125 g 206 t
ORIGIN
Query Match 12.1%; Score 135.2; DB 12; Length 635;

```
Best Local Similarity 61.58; Pred. No. 3.2e-22;
Matches 252; Conservative 0; Mismatches 153; Indels 5; Gaps 2;

Qy 607 tgtttgtcacaagcactgttcctttttgcccagatgacacccacaggagactgtttt 666
Db 3 TGAAGGTGATTAATACTGCTCACTCTCCCACTGAACATACATCATCGGGGACTGATTT 62
Qy 667 tcacactaagcactttgaggatgacacacatttataggttccatctcacaagagct 726
Db 63 TAACCATGACAACTTAAGAGATGATGTTCTGTAGGAGTCATGCTGACTACGAGACAT 122
Qy 727 acatgggtattttttacacagacaataagaggctatctcagtcgcttcacagcagc 786
Db 123 ACATGGTGATTATCTGTCAGACATCAGAGGCAATGCTTTTTTCTCATAGCATCAA-- 180
Qy 787 ctgtcccccagctccacagtgaaagagcctcccaggctatctactgtcgttgagt-- 844
Db 181 CCACCTGAGAGCCCTCCCTGAGAAAAGGGCCACATACCATCTTGCTGCTGGTGGTTT 240
Qy 845 -tttgtctcacactacactgggtgactttacgttctctctcttcttcaggaggtgacatgat 903
Db 241 CTTTGTGTCATGTAAGTGGGAGGACTTCATCATCTCAATCCACCTCAGTCCTGTTATGAT 300
Qy 904 aaatgattctctcgtagtggtgccagggttatgttgccaaatagctatgcccgaattag 963
Db 301 GTATGACCCAGTCATCTCGACTGTTTCAGAAAGTTTGTGATGAATGCCATCTACAAATTAC 360
Qy 964 tcctttgatgctaattatgctgataccaaaataattcaagactctgcaaa 1013
Db 361 TCCATTGTGTACAAATCAAGTTCTGTGTAAGCGAATAATCATGATGCTGTAATAA 410

RESULT 9
A1806860/c
LOCUS
DEFINITION wf36007.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2357653 3' similar to TR:Q62855 Q62855 PHEROMONE RECEPTOR
VN6. ; mRNA sequence.
ACCESSION A1806860
VERSION A1806860.1 GI:5393426
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 574)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 659 Std Error: 0.00
Seq primer: -40UP from Glibco
High quality sequence stop: 447.
Location/Qualifiers
1. 574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2357653"
/lab_host="DHI08"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not 1; Site.2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHR, and B-cell
NCI_CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
```

```
I.M.A.G.E. clones 297480-302087, 682632-687239,
728408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.
BASE COUNT 179 a 108 c 138 g 149 t
ORIGIN

Query Match 11.98; Score 132.2; DB 9; Length 574;
Best Local Similarity 60.48; Pred. No. 1.6e-21;
Matches 306; Conservative 0; Mismatches 193; Indels 8; Gaps 5;

Qy 514 attcttcttttctcagtggttcctcaacatgttctcatggt-gtaattcttctgtgtacac 572
Db 566 ATGTTTCTTCTTCTATGTTCTCCTCACTTCTTCAGTAGTAACAGGATCTTCTACAC 507
Qy 573 tgcggctacccccagtg-aaacggggccagtcctctgtttgttcatacagacacttcttc 631
Db 506 TGTAGCTGCTTCCAGTGTGAGCCAGACAAATGTAAGTAATAGGTAAAGTACTGCTCAGT 447
Qy 632 ttggcccatcagctacacccacagagactgttttccacactaatgactttgaggatgt 691
Db 446 TTTCTCATGAGCTTCATCTCAGGGCCACGTTCTTATTCTGACAACTTAGGGATGC 387
Qy 692 cacctttatcaggttctcaggttcctctcaagaggtcacatggtgattttttatcacagaca 751
Db 386 TTTCTGTTAGGAATACTGCTGTTCTTAATTGCATACATGCTGACTCTCTTATTTCAGGC- 326
Qy 752 ataagaggtctatctcagtgcttcacgcagcagcctgtccccc--gagctccacagtgca 809
Db 327 CTCAGAGGTGATCCCAACACCTTCACAGCAATGGTCTTCCCAAGGAGCTTCCCTCTGAGA 268
Qy 810 aaagagctccaggctattctactgctggtgagttt---tgtttcacacactcagggtgg 866
Db 267 AACGGGCCATCCAGACCATACTGCTGTGGTGAAGTTCTCTGTGGTCTCTGTGGGTGG 208
Qy 867 actttagcttctctatttcaggaggtgtgacatgataaaatgattctctctgtagtgagc 926
Db 207 ATTTTCATCATCTCTCTTCTCCTCAGCAATGTTGTGGCATATGGCCCAATCAACGACGA 148
Qy 927 tccaggttattgtggccaatagctatgccgcaatagtcctttgatgctaattatgctg 986
Db 147 TTCAGAGTCTTTGTAGTCAATGCTATGCCGTGATCAGTCCCTCTGTGCTCCTCCTCAG 88
Qy 987 ataaccataattcgaagactctgcaaa 1013
Db 87 ATAAAGAAGAACAATATTCTGAAAA 61

RESULT 10
A2553624
LOCUS
DEFINITION RPCI-23-209L1.TV RPCI-23 Mus musculus genomic clone RPCI-23-209L1,
DNA sequence.
ACCESSION A2553624
VERSION A2553624.1 GI:11233124
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 615)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
B., Levins,M., Megann,S., Tsegaye,G., Geer,K., Kroi,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-209L1.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
```

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 209 row: L column: 1
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. .615
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-209L1"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 178 a 128 c 107 g 202 t
ORIGIN

Query Match 11.6%; Score 129.4; DB 12; Length 615;
Best Local Similarity 58.3%; Pred. No. 8e-21;
Matches 284; Conservative 0; Mismatches 196; Indels 7; Gaps 3;
Qy 99 ttctgttaatttctcttaagatgacaaactctctggtttattaccatacaaatct 158
Db 80 TATGACTCAACAATGCTTTAAGATAAATTCACCTCTCACTTATGTTCTATGGAGAAT 139
Qy 159 tct-tttatccacagccagctgttggaatttcagcaaacaccatctcttctttccac 217
Db 140 GCTCTTTATATCCAAAGCTGGGTAGGAGTCTAGTCTAATATATGTCCTCTTTCTAT 199
Qy 218 atcttcacctgttttcagtcacaggtctcaagtccttaagtcattgacatgaattagtcacctg 277
Db 200 ATTTTCA---FAATCCTTAGGCGACAGACCTTAAGCCCATGGACCTTAATCTCCTGTCAACAG 256
Qy 278 tctctcatccacactactgctcttctcactcaggcaaatattggtctcttagacttcttt 337
Db 257 ACTTTTCAATACATATGCTGTTCTTCACTGCAGGATATTTTGCATACAGATATATTT 316
Qy 338 ggttcacagaatactcaggatgatcttaggtataaagtgatcttcttttaacaagggtg 397
Db 317 GAGTCAATGAATATTGAGATGACTTCAAAATGCAAAACAACTTTTACATATGCAGGTA 376
Qy 398 atgaggggctctccatctgcacccctgctctgagtgatgctccaggc---catcatc 454
Db 377 ATGAGAGGCTCTATATCTGCACCACTGCTCTGAGTGTGTTCAGGCTGTCCACATC 436
Qy 455 agccccagcatctctctctggcaagctcaaacatctctctgcaagtcacatcttagga 514
Db 437 AGTCCCAATACCTTCGCTGTGGCAAAATTTAAACATAAATAAATAACATATCA 496
Qy 515 ttcttctttctcatgggtctcaacatgttcatgttggtgtaattcttctctgctgacactg 574
Db 497 TGCCTTCTCTATATTTGGGGCTTTAAATTTGCTCTCAGTAGTAGTAACCTGATCTTCATG 556
Qy 575 cggctac 581
Db 557 TTGTGTC 563

RESULT 11
AZ222872/c

LOCUS AZ222872 610 bp DNA linear GSS 14-JUN-2000
DEFINITION RPCI-23-46M10.TJ RPCI-23 Mus musculus genomic clone RPCI-23-46M10,
DNA sequence.
ACCESSION AZ222872
VERSION AZ222872.1 GI:8530921
KEYWORDS GSS.
SOURCE house mouse
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 610)
Zhao,S., Niernan,W., Feldblyum,T., Malek,J., Shatsman,S., Aklnret
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kroi,M., de Jong,P.
and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-46M10.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 46 row: M column: 10
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. .610
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-46M10"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 194 a 114 c 124 g 178 t
ORIGIN

Query Match 11.6%; Score 128.8; DB 12; Length 610;
Best Local Similarity 59.3%; Pred. No. 1.1e-20;
Matches 273; Conservative 0; Mismatches 182; Indels 5; Gaps 3;
Qy 574 gcggctaccgccagtgaaacggggccagctctctgttgcatacagcactgttctcttt 633
Db 590 GTGGTTCCCAATGTCAGTCGAGCAACCAACAGATGCAGGTCACCTAAATCTGCTCACTCT 531
Qy 634 tgccatagctacacccacagagagactgttttccactaataagctttgagggatgca 693
Db 530 TACCCATGAACATATCATCAGAGGATGATTTTACAATATCAACCTCCAGAGATGTGT 471
Qy 694 cctttataggggttcagtgctctcctcaagagcctacatgggtgattattttatatacagacaat 753
Db 470 TTCTGTAGGAGTCATGCTGACCACAGTGTATATACATGGTCATTTAG-CTAT 412
Qy 754 aagaggtctatctcagtgcttcacgcagccagcctgtcccccagctctccacagtgaaaag 813
Db 411 CACAGGCAAGTCAAGTATCTTTATAGCATCAGCCACT-TGAGAGAGTCCCTCGAGAAAG 353

```

Qy 814 agctccaggctattcttaactggtgagc---tttgcttcacatactggtgagctt 870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 GCCACCCAGACCATCGTCGTCGGTGTCTTTCTTTGTTGTCATGACTGGGGAGACTT 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 871 tacgttctcatttcaggagdggtgacatgataaattgattctgctagtggtgctcca 930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 CATCATCTCAATCACCTACAGATATGATATGATGATGATGATGATGATGATGATGAT 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 931 ggttatgtggccaatagctatgcgcgaattagtcctttgattgataatttatgctgataa 990
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 GAAATTTATGATGATGCTTATCCACCAATTTACCCCTTTTGGTCAAAATTCAGTTCTGTAA 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 991 ccaaatattcaagctctgcgaatggttatggttttaaatat 1030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 CAGAAATATAATTATGCTGAAACCTCAATCAAAACAT 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AQL117475
LOCUS
DEFINITION HS_2187_B2_D03_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-2187 Col-6 Row-H, DNA sequence.
ACCESSION AQL117475
VERSION AQL117475.1 GI:3495266
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 467)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2187 row: H column: 6
Class: BAC ends
High quality sequence stop: 467.
FEATURES
source
1..467
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-2187 Col-6 Row-H"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 121 a 114 c 82 g 148 t 2 others
ORIGIN
Query Match 11.5%; Score 128.4; DB 12; Length 467;
Best Local Similarity 61.8%; Pred. No. 1.3e-20;
Matches 289; Conservative 0; Mismatches 168; Indels 11; Gaps 5;

Qy 381 tcttttaaacaggatgaggccctctccatctgacccccctgctctctgagtgctg 440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 TCTATNTGCACAGAGTGATGAGGGCCCTCTCCATATTCACCTGCTCTCTGAGCATGC 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 441 tccaggccat---catcagcccccagcatcttctctcttgaggcaagctcaaacatctcttg 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 TCCAGGCCATTACCATCAGCCTCAGCACCTCTCTGTTGGTTAGATTAAACATAAATTTA 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Qy 498 caagtcacatcttaggattcttctcttctcattggtcctcaacatgcttcatggt-gta 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 CAAATACGATATCCTCGCTTATTCGTTTGTGTTTGTAGCAATTTGCTTTCAGTAGTG 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 557 atctctgctgtacactgcggctaccgccagtgaaacggggccagctctctctgtttgtcat 616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 ACATGATAATCTACACTGTAGGTTATTCCAATGACCCAG----ATAATTTTGAATATCAG 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 617 acagcaactgttctctttttgccccatgagctacacccacacagagagactgtttttcacacactaa 676
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 CAAATATTTGCCACATTTTTCCTCAATGAATGCTCTCATCAGGACGCTATTTCTTATGCTGTG 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 677 gacttgaggatgtcacctttatagggctcatggtctctctcaagagagctacatgggtgat 736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 ATTATCCAGAGATGCCCTTCTTCATAGGAATCACCGCTGCTCTCAAGTGTATACATGGTGCAT 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 737 tattttacagacaataagaggctatctcagtccttcacgcagccagcctgtctcccca 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 TCTTTTGTCCAGGC-ATCAGAGGCACCTCCAGCAGCTTTTCACAGCAGCAGCCTTATATTAA 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 797 g--tctcaccagtaaaagagcctcccgagctatcttactgctggtga 842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 GGACTTCTCTAGTGAATAATGGCCACCAAGACCATCTCTGATGCTGTGTA 466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AZ810472
LOCUS
DEFINITION AZ810472 563 bp DNA linear GSS 20-FEB-2001
clone UUGC2M0076E06 F, DNA sequence.
ACCESSION AZ810472
VERSION AZ810472.1 GI:12977756
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 563)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0076 row: E column: 06
Seq primer: CGTTGTAACAGCAGGCCAGT
Class: plasmid ends
High quality sequence stop: 563.
FEATURES
source
1..563
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0076E06"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA

```

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match	11.4%	Score 127.4	DB 12	Length 563
Best Local Similarity	61.4%	Pred. No. 2.4e-20		
Matches 256	Conservative 0	Mismatches 156	Indels 5	Gaps 3
QY	613	tcatacagcactgtctcttttgcacatgagctacacccacagagagctgtttttcacac	672	
Db				
Db	9	TCACATAATCCTGCTTACTCTTCGGCATGAACTCCATCATCAGGGTATTCAATTTTACAG	68	
QY	673	taatgaccttggaggatgtcacctttatagggtttcatgtgtctctcacaagaggtacatatgg	732	
Db				
Db	69	TGCAACCTCCAGAGATGTATTCTTGTAGAAATGATGCTAAACCACAAAGTACATACATGG	128	
QY	733	tgattattttatcacacaataagaggctatctcagtcgcttcacgcagcagcagcctgtcc	792	
Db				
Db	129	TGATTATCTTATGTAGAC - ATCAGAGGCGATGCAAGCACCTTCATAGCATCAAGC - ACCT	186	
QY	793	ccgagctccacagtgaaaagagcctcccaggtctattactgctggtgagct-----tttgt	849	
Db				
Db	187	GAGAGCATCCCTTGAGAAAAGGCCACCCAGACTATCTTGCTGCTAGTGATTTTCTTTGT	246	
QY	850	cttcacataactgggtggacttttaagttctcatcttcaggagggtgtgacatggataataatga	909	
Db				
Db	247	GGTCATGTACTGGGTGGAGTTCATCATCTCAACCATATCATGTTCTGTGTGGAGGTATGA	306	
QY	910	tctctgctagttggtgctccaggttatgtggccaatagctatgcccgaattagtccttt	969	
Db				
Db	307	CACAGTCATCCTGAGGGTTCAGAGTTTTTGATGTATGCCCTATCCACAAATTAGTCCTTT	366	
QY	970	gaigtctaattatgctgataaccaaataattcaagactctgcgaatgttatggttaa	1026	
Db				
Db	367	GGTACAAATCAGTCTCATACAGAAATAATAATATGGTGAACCAACATGTACTCAA	423	

RESULT	14
AZ533591	
LOCUS	638 bp DNA linear GSS 20-NOV-2000
DEFINITION	Mus musculus genomic clone RPCI-23-209J1,
ACCESSION	AZ533591
VERSION	AZ533591.1
KEYWORDS	GI:112333013
SOURCE	GSS.
ORGANISM	house mouse.
	Mus musculus.
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
REFERENCE	1 (bases 1 to 638)
AUTHORS	Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret B., Levins,M., Megann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE	Mouse BAC End Sequences from Library RPCI-23
JOURNAL	Unpublished (1999)
COMMENT	Other GSSs: RPCI-23-209J1.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA

```

Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 209 row: J column: 1
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. .638
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-209J1"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
184 a 135 c 105 g 214 t

```

Query Match	11.4%;	Score 127;	DB 12;	Length 638;
Best Local Similarity	59.3%;	Pred. No. 3.le=20;		
Matches	272;	Conservative	0;	Mismatches 180;
				Indels 7; Gaps 3;
QY	99	tttgctttaattctcttaagatgatcaaaacttcctcggtttattaccatcaaaatct	158	
DB	39	TATGACTCAACATGCTCTTAAGATAAATTCACCTCTCAGTTATGTTCTCATTTGGAGAAT	98	
QY	159	tct-tttatocacagccagcttggaaatttcagcaaacaccatcctctcttctttccac	217	
DB	99	GCCTTTATATCCAGCTGGGTTAGGAGTCTAGCTAAATATGTCCTCTGTTTCTTAT	158	
QY	218	atcttcaccttgttttcagtcagaggtctaaagtcattgacatgataattagtcacctg	277	
DB	159	ATTTTCA---TAATCTAGGGCACACACTTAAGCCCATGACCTAATCTCCTGTCAACAG	215	
QY	278	tctctatocacatctgtctcttcttcactcaggaatattggtgtccttagactcttt	337	
DB	216	ACTTTTCATTACATAATGCTGTCTTTCACATGACGAGAGATATTTGTCATACAGATATATTT	275	
QY	338	gggtcacagaatactcagatgactctaggataaaggtcattgtctcttttaaacaaagtg	397	
DB	276	GAGTCAATGNAATATTGAGAATGACTTCAAATGCAAAACAACCTTTTACATATGCGAGGTA	335	
QY	398	atgaggggcctctccatctgcaacccctgcctcctgagtgctccaggc---catcatc	454	
DB	336	ATGAGAGGCGCTCTATCTGCACCACTGCTGCTGAGTGTGTCAGGCGTGCACCATC	395	
QY	455	agccccagcatctctccctggcaagctcaaacctctctcgactcaggtcactcttagga	514	
DB	396	AGTCCCAATACCTTCGCTGTGTGGCAAAATTTAAACATAAACAATAAANAATACATATCAAT	455	
QY	515	tctctcttcttcacgtgggtccotcaacatgcttcattgggt	553	
DB	456	GCCTTCTCTATATTTTGGTCTTTTAAATTTTCTCTCTTCAGT	494	

[illegible]

CH230-82L13, DNA sequence.

BH267496

BH267496.1 GI:17179806

GSS.

ORGANISM

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 740)

Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn

,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de

Jong,P. and Fraser,C.M.

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

Unpublished (1999)

Other_GSSs: CH230-82L13.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(<http://www.chori.org/bacpac/rat230.htm>). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/orering_information.htm). BAC end

page: http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html

Plate: 82 row: L column: 13

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..740

/organism="Rattus norvegicus"

/strain="BN/SSHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-82L13"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pFARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;

CHORI-230 Rat (BN/SSHsd/MCW) BAC library produced by

Pieter de Jong"

BASE COUNT 250 a 136 c 136 g 218 t

ORIGIN

Query Match 11.4%; Score 126.8; DB 12; Length 740;
Best Local Similarity 64.6%; Pred. No. 3.7e-20;
Matches 252; Conservative 0; Mismatches 132; Indels 6; Gaps 4;

Qy 627 tccttttgcctagctacacccagaggactgttttcacactaatgactttgagg 686

Db 740 TCACCTCTTTCCCATGAACATCATCATCAGGGGACTGATTTTAAACAATGACAACTCCAGA 681

Qy 687 gatgtcaacctttatagggttcattcctctcaagaggctacatgggtattttataac 746

Db 680 GATGCTTTTCTA-GTAGGAGTCTGCTGACAAACAAGCATGTACATGGTGATTCTTGTC 622

Qy 747 agacaataagaggctatctcagtccttcacgcagcccgctcccgagctccaccag 806

Db 621 AGAC-ATCAGAGGAGTGCAAGCATCTTCATPAGCATAGTCACCT-TGAGAGCATCCCCCTG 564

Qy 807 tgaagaagcctccaggctacttactgtgtgtgag---ttttgtcttcacatactggg 863

Db 563 AGAAAAGGCCACCCAGACCATCTTGCTGTTGGTGGGTTTTTTTGTGCTCATGTACTGGG 504

Qy 864 tggactttacgttctcattttcagagggtgtgacatggaataaatgattctctgctagtgt 923

Db 503 TGGACTTCATCATCTCATCCACCTCAGTCTTATTATGGATGTATGACCCAGTCATCCTGA 444

Qy 924 ggctccagggttatgtggccaatagctatgccgaatagctcctttgtatgctaatattatg 983

Db 443 CTGTTCAAGATTGTGTGATGATGCCTATCCCACAATCTCTTTGATACAAATCAGTT 384

Qy 984 ctgataacacaaatattcaagactctgcaaa 1013

Db 383 CTGATAACCGAATAATCATTACGCTGAAAA 354

Search completed: July 21, 2002, 02:40:43
Job time: 11693 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 01:55:45 ; Search time 376.86 Seconds
(without alignments)
5075.207 Million cell updates/sec

Title: US-09-728-309-1
Perfect score: 1114
Sequence: 1 attccagatcatagagatgt.....gagaggttaattcattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	231.4	20.8	1331	18 AAT69546	Rat pheromone rece
2	226.4	20.3	1538	18 AAT69549	Rat pheromone rece
3	221.8	19.9	1386	18 AAT69545	Rat pheromone rece
4	215	19.3	584	22 ABA60235	Human foetal liver
5	215	19.3	584	22 ABA28536	Probe #7002 for ge
6	215	19.3	584	22 AAK08511	Human brain expres
7	215	19.3	584	22 AAK34394	Human bone marrow
8	215	19.3	584	22 AAI40117	Probe #8803 used t
9	207.4	18.6	1496	18 AAT69547	Rat pheromone rece

10	205.8	18.5	1053	18 AAT69548	Rat pheromone rece
11	204.4	18.3	1265	18 AAT69550	Rat pheromone rece
12	184.6	16.6	1313	24 AAS98047	Human DNA for pote
13	181.6	16.3	1015	24 AAD24968	Human G-protein co
14	163.2	14.6	628	22 AAH74904	Probe used to Isol
15	158.4	14.2	632	22 AAH74901	Probe used to Isol
16	152.4	13.7	461	22 AAK06912	Human brain expres
17	152.4	13.7	461	22 AAK32640	Human bone marrow
18	152.4	13.7	461	22 AAI38455	Probe #7141 used t
19	151	13.6	632	22 AAH74899	Probe used to Isol
20	149.4	13.4	628	22 AAH74900	Probe used to Isol
21	143.8	12.9	640	22 AAS63212	Human purified sec
22	143.4	12.9	653	22 AAH74903	Probe used to Isol
23	138	12.4	662	22 AAH74902	Probe used to Isol
24	135.4	12.2	636	22 AAH74905	Probe used to Isol
25	134.8	12.1	300	22 ABA72774	Human foetal liver
26	134.8	12.1	300	22 ABA38413	Probe #16879 for g
27	134.8	12.1	300	22 AAK21203	Human brain expres
28	134.8	12.1	300	22 AAK47360	Human bone marrow
29	134.8	12.1	300	22 AAI53196	Probe #21882 used
30	132.2	11.9	574	24 AAS98151	Human DNA for pote
31	128.6	11.5	512	24 ABA05725	Rat pheromone rece
32	127	11.4	512	24 ABA05712	Rat pheromone rece
33	127	11.4	512	24 ABA05731	Rat pheromone rece
34	125.4	11.3	512	24 ABA05704	Rat pheromone rece
35	125.4	11.3	512	24 ABA05705	Rat pheromone rece
36	125.4	11.3	512	24 ABA05710	Rat pheromone rece
37	124.2	11.1	512	24 ABA05721	Rat pheromone rece
38	123.6	11.1	530	18 AAT69551	Human pheromone re
39	120.8	10.8	512	24 ABA05703	Rat pheromone rece
40	120.8	10.8	512	24 ABA05708	Rat pheromone rece
41	119.2	10.7	512	24 ABA05701	Rat pheromone rece
42	119.2	10.7	512	24 ABA05702	Rat pheromone rece
43	119.2	10.7	512	24 ABA05711	Rat pheromone rece
44	119.2	10.7	512	24 ABA05714	Rat pheromone rece
45	119.2	10.7	512	24 ABA05715	Rat pheromone rece

ALIGNMENTS

RESULT	1
AAT69546	
ID	AAT69546 standard; DNA; 1331 BP.
AC	AAT69546;
XX	
DF	26-AUG-1997 (first entry)
XX	
DE	Rat pheromone receptor VN3 cDNA.
XX	
KW	Pheromone receptor; vomeronasal sensory neuron; social behaviour;
KW	maternal behaviour; reproductive behaviour; fertility;
KW	hormone secretion; ss.
XX	
OS	Rattus sp.
XX	
PH	Key
CDS	Location/Qualifiers
FT	207..1142
FT	/*tag= a
FT	1038..1040
FT	/*tag= b
FT	/transl_except= pos:1038..1040:_aa:Gly
FT	/note= "GTA codes for valine"
FT	1086..1088
FT	/*tag= c
FT	/transl_except= pos:1086..1088:_aa:Glu
FT	/note= "GGA codes for glycine"
XX	
WO	WO9714790-A1.
PD	24-APR-1997.
XX	

receptors VN3, VN4, VN5, VN6 and VN7 (AAW19104-08), members of a novel family of presumed 7-transmembrane domain receptors that are evolutionarily independent of the odorant receptors of the main olfactory epithelium. These clones, and a clone for VN2 (sequence not provided), were isolated from rat vomeronasal organ cDNA libraries by PCR and hybridisation. A differential cloning method was used to isolate VN7 cDNA (AA169545). VN1-VN7 (not VN2) clones have been deposited in pBluescript as ATCC 97294-97299. A human homologue, HG25 (AA069551), has also been isolated. VN polypeptides can be expressed in host-vector systems for use in identifying modulators for control of maternal, reproductive and social behaviour, to increase fertility, control hormone secretion and to regulate food uptake in humans and animals.

Qy	892	tgtagacatggataaatagtctctgctagtggtccaggttattgtggccaatagcta	951
Db	1034	tatg---cacacggtgaaccaatccacacagtatctcagatgatggtctccaatagcta	1090
Qy	952	tgcgcgaattagccctttgatgctaatctatgctgataacaaatattcaagactctgca	1011
Db	1091	tgcacccctcagccctttgctgttaattgcttactgtaataatgaattagtaggttttgaa	1150
RESULT 3			
AAT69545			
ID	AAT69545 standard; DNA; 1386 BP.		
XX			
AC	AAT69545;		
XX			
DT	26-AUG-1997 (first entry)		
XX			
DE	Rat pheromone receptor VN1 cDNA.		
XX			
KW	Pheromone receptor; vomeronasal sensory neuron; social behaviour;		
KW	maternal behaviour; reproductive behaviour; fertility;		
KW	hormone secretion; ss.		
XX			
OS	Rattus sp.		
XX			
PH	Key	Location/Qualifiers	
FT	CDS	270..1217	
FT		,'*tag- a	
XX			
PN	WO9714790-A1.		
XX			
PD	24-APR-1997.		
XX			
XX	18-OCT-1996;	96WO-US16637.	
PF			
XX			
PR	19-OCT-1995;	95US-0005698.	
XX			
PA	(UYCO) UNIV COLUMBIA NEW YORK.		
XX			
PI	Axel R, Dulac C;		
XX			
DR	WFI; 1997-245107/22.		
DR	P-PSDB; AAW19103.		
XX			
PT	Nucleic acid molecule encoding vertebrate pheromone receptor -		
PT	useful to identify modulators for control of reproductive and social		
PT	behaviour, fertility and hormone secretion		
XX			
PS	Claim 1; Fig 8; 123pp; English.		
XX			
CC	A cDNA clone (AAT69545) codes for rat pheromone receptor VN1		
CC	(AAW19103), a member of a novel family of presumed 7-transmembrane		
CC	domain receptors that are evolutionary independent of the odorant		
CC	receptors of the main olfactory epithelium (MOE). It was isolated		
CC	by differential screening of cDNA libraries constructed from rat		
CC	single vomeronasal sensory neurons. A VN1 insert in pBluescript		
CC	has been deposited as ATCC 97294. VN1 is expressed only in about		
CC	4% of VSN and never in the MOE. cDNA clones for VN2-7 were obtd.		
CC	by PCR and hybridisation (see also AAT69546-50, VN2 sequence not		
CC	given). VN polypeptides (AAW19103-09) can be expressed in host-		
CC	vector systems for use in identifying modulators for control of		
CC	maternal, reproductive and social behavior, to increase fertility,		
CC	control hormone secretion and to regulate food uptake in humans and		
CC	animals. A human homologue, HG25 (AAT69551), has also been obtd.		
XX			
SQ	Sequence 1386 BP; 357 A; 322 C; 271 G; 435 T; 1 other:		

Db	278	taagaaacgacgactacacgtattctaataaggaatacctttttgcgtgaattgg	337
Qy	178	cttttggaatttcagaaacacacactctctcttttccacatctccactttggtttcaag	237
Db	338	catgtgagctcagcccaatagcctcctactctctccacatctccaagttaattttgg	397
Qy	238	tcacagggtcaagtcattgacatgaataatagtcacctgtctctccacatactgct	297
Db	398	gcagagggtccagactcactgacctgccattggtctcttgccctaatacaacttactat	457
Qy	298	gctcttcaactcagcaatattggtgtccttagactctttgtttgcacagaaatactcagga	357
Db	458	gctactgtagcggc--attcatgacacagacactttattcttggagagggtggga	514
Qy	358	tgacttaggtataaggtcattgtcttttaaacagaggatgaggggcctctccatctg	417
Db	515	tgacataatgtataacccctctctactctgtacagaacttttagaggctctctcttgg	574
Qy	418	cacccctgcctcctgagtggtctccaggccatca---tcagcccagcatcttctctct	474
Db	575	taccagctcgtgtgagtgctcgtcaggccatcctcctcagtcaccagaagctcgtgtt	634
Qy	475	ggcaagctcaaacatcctctctgcaggtcacatcttagagatctctcttttctcatgggt	534
Db	635	agcaaggtcaaacataagccttcccatctccatctcctgtgcactcttctctctgagtg	694
Qy	535	ctccaacatgttcaatt--ggtgttaactctctgctacactgcggctacccccagtgaaac	593
Db	695	cctctacatgttcatagcagtcactcttagtatccatcattgcacccccaaattgac	754
Qy	594	ggggccagtc--ttctgtttgtcatacagcactgttctcttttgcccatgagctacaccca	652
Db	755	cacgaatgactttattcatgttactcagtggtgctctattctcaaccagtgattacctcat	814
Qy	653	cagagactgttttcaacaaataagctttgaggatgtcaacctttatagggttoatgggt	712
Db	815	gcaagcatgttttctacactgctggccatcaggatgtctctcttatagtctcatggt	874
Qy	713	ccctcagaaggctacatgggtattattttaacagacaaataagaggtctactcagtgcc	772
Db	875	cctgtcaacatggtacatgggtgctctctgtgtaggc--acaaggaaacagaccggcatc	933
Qy	773	ttcacgcagccagcctgtcccc--gagtcctaccagtgaaaagagcctcccaggtctatct	830
Db	934	ttcaggttacaccctttccccaaagcatccccagacaaacaaaggccaccggttccatcc	993
Qy	831	tactgctgggtgagtttgtctccataactcgggtggagcttttaactgttctcattttcaggag	890
Db	994	tgatgctcagacttatttgttctgtatgtctgtcttttgacagcatgtctctgcagctcaa	1053
Qy	891	gtgtgacatggataaataattctctgtagtgggtctccaggttatttgggccaatagct	950
Db	1054	gaactagtactgtagatccaatatctattcttatcaactatttatggtgcacatct	1113
Qy	951	atgccgcaattagctcttttgatgctaattattgtgtgataacccaataatccaagactctg	1009
Db	1114	atgccacagtaagccctttgtgtattgttctgcactgaaaacatatagtttaactctttg	1172

RESULT 4

ABA60235

ABA60235
ID ABA60235 standard; DNA: 584 BP.

XX
XX

AC ABA60235;

XX

DT 01-FEB-2002 (first entry)

[illegible]

DE Human foetal liver single exon nucleic acid probe #8540.

XX

human; foetal liver; gene expression; single exon nucleotides

[illegible]

Q 112

Query Match 19.3% Score 215 DB 22 Length 584

Best Local Similarity 64.78; Pred. No. 3e-50;

QY 190 aqcaaacaccatccttcttcttcttcttccacacatcttcacacttgcactttgcaggtcacaggtctaa 249

[illegible]

Db
4 agccaacttcctctattcttccaaatcttctactccttcaggatcacaggcctaa 63

[illegible]

Qy 250 gtccattgacatgataattagtcacctgtctctcatccacatactgctgctcttca

64 db

50 5' yccccaclyaccccccatcacctgtcacctggcccttgtccacttaggggatgctccctcattgt 123

Qy 310 ggcaatatggtgtcccttagactctctttggttcacagaaatactcaggatgattcttagata 369

[illegible]

Db 124 ggtctcttggcatctccagacctgttgagtcactgtattttcagaatgacttcaagtg 183

[illegible]

QY 3/0 taagggtcattgtctctttttaacaagggtgatgaggcctctccatctgcacccccctgct 429

[illegible]

184 caaaggcattctctgcatgcacaggggtgatgaggagcctctccatctgcaccacctgcct 243

Qy 430 cctgaqtatqctccaggc---catcatcagccccagcatcttctccttgcacaaagctcaa 486

Db 424 gctaatgataactgatacatttcttccataagctacattatcagagcctgtc 483
Qy 665 ttccacactaatgactttgagggtgacacctttataggggttcattggtctctcctcaagagg 724
Db 484 ttctatgctgcactgcttacgaatgctcttctttagcaatcagcagcctcaagtgc 543
Qy 725 ctacatggtgattttttatcacagaca 751
Db 544 atacatggttaattctctgttcaggca 570

RESULT 5
ABA28536
ID ABA28536 standard; DNA; 584 BP.
XX
AC ABA28536;
DT 23-JAN-2002 (first entry)
DE Probe #7002 for gene expression analysis in human heart cell sample.
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
XX WO200157274-A2.
PN Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX 09-AUG-2001.
PD
PF 30-JAN-2001; 2001WO-US00666.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-488899/53.
DR
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
XX
XX Claim 1; SEQ ID No 7002; 530pp; English.
PS
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 584 BP; 127 A; 173 C; 107 G; 177 T; 0 other;

Query Match 19.38; Score 215; DB 22; Length 584;
Best Local Similarity 64.7%; Pred. No. 3e-50;
Matches 367; Conservative 0; Mismatches 195; Indels 5; Gaps 3;
Qy 190 agcaaacaccatctcttcttccacatcttcacctttgttttcagtcacaggtctaa 249

Db 4 agccaaacacttctctctatttctccaaatcttctcactcttcagatcacaggcctaa 63
Qy 250 gtccattgacatgataaattagtcacctgtctctccacacatactgtctctctcactca 309
Db 64 gccacatgaccccatcacctgtcacctggcccttgctccacttaggagtgctctcattgt 123
Qy 310 ggaatatgtgtctcttagactctcttggttccagaaatactcagatgatcttaggta 369
Db 124 ggtcttctggcatctccagacctgtttagtcactgtattttcagaatgacttcaagt 183
Qy 370 taaggtcattgtcttttaaacaaagggtgaggggctctccatctgcacccctgcct 429
Db 184 taaggcattctctgcatgcacaggggtgatggaggcctctccatctgcacccctgcct 243
Qy 430 cctgagtggtctccaggc---catcatcagccccagcatcttctccttggcacaagctcaa 486
Db 244 cctgagcatgctccaggctgtcgccatcagccctggcacctcctggtctgcaagattaa 303
Qy 487 acatcctctgcaagtccacatctcttaggattctctcttctctcattggtctcacaatgt- 545
Db 304 acagaaattcaagggttacatcttctcactctctctctctctatgggtctcagttgtc 363
Qy 546 tcattggttaattctctgtgtacactgctggtcaccctcagtgaaacggggccagctct 605
Db 364 tctcagtagtaacctgtctctcactgtggtcttcttaattgggacacagcgtgtgt 423
Qy 606 -ctgttctcatcacgacactgttctctttttgtcccatgagctacacccacacagagactgtt 664
Db 424 gctaaatcatcagtaataactgacactctctctctccataagctacattatcagaggcctgtc 483
Qy 665 ttccacactaatgactttgagggtgacacctttataggggttcattggtctctcctcaagagg 724
Db 484 ttctatgctgcactgcttacgaatgctcttctttagcaatcagcagcctcaagtgc 543
Qy 725 ctacatggtgattttttatcacagaca 751
Db 544 atacatggttaattctctgttcaggca 570

RESULT 6
AAK08511
ID AAK08511 standard; DNA; 584 BP.
XX
AC AAK08511;
DT 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe SEQ ID NO: 8502.
DE
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
PN
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI

```

XX DR WPI; 2001-483446/52.
XX KW Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488900/53.
XX DR
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 8951; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX CC the probes of the invention.
XX SQ Sequence 584 BP; 127 A; 173 C; 107 G; 177 T; 0 other;

Query Match 19.3%; Score 215; DB 22; Length 584;
Best Local Similarity 64.7%; Pred. No. 3e-50;
Matches 367; Conservative 0; Mismatches 195; Indels 5; Gaps 3;

QY 190 agcaaacaccatcctcttcttccacatcttcacacctttgttttcagtcacaggtctaa 249
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4 agccaacaactctctctattcttccaaatcttctcactctctcaggtacacaggtctaa 63

QY 250 gtccattgacatgaattagtcacctgtctctccatccacatactgtctcttcaactca 309
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 gccactgacccatcacctgtcacctggccctgtccacttagggatgtctctcattgt 123

QY 310 ggcaaatattggtgctcttagactctttgttccacagaatactcaggtatcttagtta 369
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
124 ggtctcttgcatctccacagacctgttgagtcactgtatttcagaatgacttcaagt 183

QY 370 taaggctattgtctttttaacaagggtgtaggggctctccatctgcacccctgcct 429
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 taaggcatctcttgcatgcacaggggtgtagggagcctctccatctgcacccctgcct 243

QY 430 cctgagtgtctccaggc---catcatcagcccccagcatcttctcttggcagaagctcaa 486
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 cctgagcatgtctccagggtgtgccatcagccctggcaacctctggtctgcaaggattaa 303

QY 487 acatctcttgcaagtcacatcttaggtattctctcttctcttctcttctcaactgt- 545
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 acagaattcaaaaggttacatcttctcactcttctctctctctctctctctctctctctc 363

QY 546 tcattggttaattctctgtgtacactgcggtacccccagtgaaacggggcagctctt 605
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 tctcagtagtaacctgtctctccactgtggcttcttcttaagtacattatcaggagcgtgc 423

QY 606 -ctgtttgtacacagcactgtctctttttggccatgagctacacacacagggactgtt 664
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 gctaaatcagtaataactcgtacactcttctctccatgaagctacattatcaggagcgtgc 483

QY 665 ttccactaatgacttctgagggtgtccaccttttatagggttctctctctcaagagg 724
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 ttctatgctgccactgtctacgaatgtctcttctgttagcaatcatgcagccctcaagtgc 543

QY 725 ctacatgggtattattttatcacagaca 751
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
544 atacatggtaattctcttcttcaggca 570

RESULT 7
AAK34394
ID AAK34394 standard; DNA; 584 BP.
XX AC
XX AAK34394;
XX DT
XX 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 8951.

```

```

XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488900/53.
XX DR
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 8951; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX CC the probes of the invention.
XX SQ Sequence 584 BP; 127 A; 173 C; 107 G; 177 T; 0 other;

Query Match 19.3%; Score 215; DB 22; Length 584;
Best Local Similarity 64.7%; Pred. No. 3e-50;
Matches 367; Conservative 0; Mismatches 195; Indels 5; Gaps 3;

QY 190 agcaaacaccatcctcttcttccacatcttcacacctttgttttcagtcacaggtctaa 249
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4 agccaacaactctctctattcttccaaatcttctcactctctcaggtacacaggtctaa 63

QY 250 gtccattgacatgaattagtcacctgtctctccatccacatactgtctcttcaactca 309
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 gccactgacccatcacctgtcacctggccctgtccacttagggatgtctctcattgt 123

QY 310 ggcaaatattggtgctcttagactctttgttccacagaatactcaggtatcttagtta 369
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
124 ggtctcttgcatctccacagacctgttgagtcactgtatttcagaatgacttcaagt 183

QY 370 taaggctattgtctttttaacaagggtgtaggggctctccatctgcacccctgcct 429
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 taaggcatctcttgcatgcacaggggtgtagggagcctctccatctgcacccctgcct 243

QY 430 cctgagtgtctccaggc---catcatcagcccccagcatcttctcttggcagaagctcaa 486
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 cctgagcatgtctccagggtgtgccatcagccctggcaacctctggtctgcaaggattaa 303

QY 487 acatctcttgcaagtcacatcttaggtattctctcttctctctctctcactcaactgt- 545
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 acagaattcaaaaggttacatcttctcactcttctctctctctctctctctctctctc 363

QY 546 tcattggttaattctctgtgtacactgcggtacccccagtgaaacggggcagctctt 605
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 tctcagtagtaacctgtctctccactgtggcttcttcttaagtacattatcaggagcgtgc 423

QY 606 -ctgtttgtacacagcactgtctctttttggccatgagctacacacacagggactgtt 664
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 gctaaatcagtaataactcgtacactcttctctccatgaagctacattatcaggagcgtgc 483

QY 665 ttccactaatgacttctgagggtgtccaccttttatagggttctctctctcaagagg 724
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 ttctatgctgccactgtctacgaatgtctcttctgttagcaatcatgcagccctcaagtgc 543

QY 725 ctacatgggtattattttatcacagaca 751
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
544 atacatggtaattctcttcttcaggca 570

RESULT 7
AAK34394
ID AAK34394 standard; DNA; 584 BP.
XX AC
XX AAK34394;
XX DT
XX 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 8951.

```

QY 606 -ctgtttgtcatcacagcactgttctcttttttggccatgagctacaccacaggagactgtt 664
 Db 424 gtaagatcatcagtaaatcatcagcacttcttccataagctacattatcaggagctgtc 483
 QY 665 ttccacactaatgactttgaggatgtcacctttataggggttcattgctctccaagagg 724
 Db 484 ttccatgctccactgcttacgaatgtctcttctttagcaaatcatgcagccctcaagtgc 543
 QY 725 ctacatggtgattattttatcacagaca 751
 Db 544 atacatggttaattctctgttcaggca 570

RESULT 8
 AAI40117
 ID AAI40117 standard; DNA; 584 BP.
 XX
 AC AAI40117;
 DT 17-OCT-2001 (first entry)
 DE Probe #8803 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX Homo sapiens.
 OS
 PN W0200157272-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234587.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-488897/53.
 DR Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 8803; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 XX Sequence 584 BP; 127 A; 173 C; 107 G; 177 T; 0 other;

Query Match 19.3%; Score 215; DB 22; Length 584;
 Best Local Similarity 64.7%; Pred. No. 3e-50;
 Matches 367; Conservative 0; Mismatches 195; Indels 5; Gaps 3;

QY 190 agcaaacaccatctcttcttttccacatcttcacacttctgttttcagtcacaggtctaa 249
 Db 4 agccaaacactctctcttcttccaaatctcttctcactctcttcaggtacacaggcctaa 63
 QY 250 gtccattgacatgataattgacactgtctctccacacatcactgctgtcttccactca 309
 Db 64 gccactgaccccatcacctgtcacctggcccttgtccacttagggatgctctctattgt 123

QY 310 ggaatatattggtgtccttagactcttcttgggttcacagaaatactcagagatgatcttaggta 369
 Db 124 ggtctcttggcatctccagacactgttttgagtcactgtattttcagaatgacttcaagtgc 183
 QY 370 taaggctcattgtctttttaaacaagtgatgaggccctctccatctgcacccctgcct 429
 Db 184 taaggctattctctgcatgcacaggggtgatggaggcctctctccatctgcacccctgcct 243
 QY 430 cctgagtggtctccagc---catcatcagcccccagcactctctcttcttggcgaagctcaa 486
 Db 244 cctgagcatgctccaggctgtcgcacatcagccctggcacctcctggtctgcaaggattaa 303
 QY 487 acatcctcttgcgaagtcaacatctcttaggattctctcttcttctctatggtctcacaatgt- 545
 Db 304 acagaaattccaagggttacatcttctcactctctctctctctctcttctctctctggttgc 363
 QY 546 tcattggtgtaattctctgtgtacactggtgtacacccagtgaaacggggccagctctt 605
 Db 364 tctcagtagtaacctgctctctccactgtggtcttcttaataaggacacagcgggtgt 423
 QY 606 -ctgtttgtcatcacagcactgttctcttttggcccatgagctacacccacagggagactgtt 664
 Db 424 gctaaagtatcagtaaatcactgatacttcttctccataagctacattatcaggagcctgtc 483
 QY 665 ttccacactaatgactttgaggatgtcacctttataggttcatggtctctccaagagg 724
 Db 484 ttccatgctccactgcttaccgaatgtcttcttctttagcaatcatgcagccctcaagtgc 543
 QY 725 ctacatggtgattattttatcacagaca 751
 Db 544 atacatggttaattctctgttcaggca 570

RESULT 9
 AAT69547
 ID AAT69547 standard; DNA; 1496 BP.
 XX
 AC AAT69547;
 XX
 DT 26-AUG-1997 (first entry)
 DE Rat pheromone receptor VN4 cDNA.
 XX
 KW Pheromone receptor; vomeronasal sensory neuron; social behaviour;
 KW maternal behaviour; reproductive behaviour; fertility;
 KW hormone secretion; ss.
 XX
 OS Rattus sp.
 XX
 PH Key Location/Qualifiers
 FT CDS 351..1283
 FT /*tag= a
 XX
 XX W09714790-A1.
 XX
 XX 24-APR-1997.
 XX
 XX 18-OCT-1996; 96WO-US16637.
 XX
 XX 19-OCT-1995; 95US-0005698.
 XX
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA
 XX Axel R, Dulac C;
 XX
 XX WPI; 1997-245107/22.
 DR P-PSDB; AAW19105.
 XX
 XX Nucleic acid molecule encoding vertebrate pheromone receptor -
 PT useful to identify modulators for control of reproductive and social
 PT behaviour, fertility and hormone secretion
 XX


```

Matches 470; Conservative 0; Mismatches 322; Indels 11; Gaps 6;
QY 207 ttctttccacattccaccttggttttctcagtcacaggttaagtcattgacatgataa 266
Db 1 tttttccacattctcattgctctcttgaaagacagatcagccattgactctcaca 60
QY 267 ttagtccactgtctcattccacactactgctgtcttccactcaggcaaatattggtgtcct 326
Db 61 ttctttcttattcttaaccacaaatagctgttataactattggacttatagctgc-- 118
QY 327 tagactcttttggttccacagaaactaccagatgatctcttaggtataaggtcatgtctctt 386
Db 119 -agacatgtttatgtctcggggagatggatctctaccacatgccagtccttatctatt 177
QY 387 taacaaaggtgataggggctctccatctgcacccccctgcctcctcgtgagtggtctccagg 446
Db 178 tggacaggtttttaggggttttacccttggctactcgtctcgtgaaagtctcttggga 237
QY 447 ceatca---tcagccccagcatcttctccttggcgaagctcaaacatcctctgcgaagtc 503
Db 238 ccactcactcagctcagagctcctgtttaaacaactttaacaataaaatcctcccatc 297
QY 504 acatcttaggatctctcttcttctcaggtcctcaacatgttcattgg-tgtaaatcttc 562
Db 298 acatctcaggtgccttcttcttctgtgtctctatatatttttggcagtcacctct 357
QY 563 tctgttacactcggctacccccagtgaaac-ggggcagctctctgttctgttcatacagc 621
Db 358 ttttaacaactgtaccccccaattgttacttcacagataatttatgtgtactaaat 417
QY 622 actgtctcttttggccatgagctacaccacacagagagactgtttttcacactaatgactt 681
Db 418 cctgttactttaccactatgatttactccagaacaagcatgttttccacaccatggcca 477
QY 682 tggaggtatcactttataggtttcattgctctcctcaagaggtcactatggtgattatt 741
Db 478 tcaggggaagccctcttattggtctcattggtcctgctccagtggtgactatggtgtcttc 537
QY 742 tatacagacaataaaggctatctcagtcgcttcacagcagcagcctgtc---cccagagtc 799
Db 538 tatggagag-acaaagatcaggccggcagcttctcagcagccagccttcttcaaaagt 596
QY 800 tcaccagtgaagagcctccaggtctattctactgctggtgagtttctgtctcacatc 859
Db 597 tccccagagcaaaaggccacagaccatcatgattctcatgagctcttctgtggtctc 656
QY 860 tgggtgactttacgttctcatttccagggaggtgtgacatggataaaatgattctctgcta 919
Db 657 tacattttggaaaaatgtgtctctactcttaggatgacattcaaggatgggtcgaatgttc 716
QY 920 gtgtggtccaggttatttggccaatagctatgcgcgaattagctccttctgactaatt 979
Db 717 tactgtgtccaaattattgtgtcccatagctatgcccaccatcagcccttttgtgttatt 776
QY 980 tatgtctgataaccacaatatcaaa 1002
Db 777 tgcacagaaaagcgtataataa 799

RESULT 11
AAT69550
ID AAT69550 standard; DNA; 1265 BP.
XX
AC AAT69550;
XX
DT 25-AUG-1997 (first entry)
XX
DE Rat pheromone receptor VN7 cDNA.
XX
KW Pheromone receptor; vomeronasal sensory neuron; social behaviour;
KW maternal behaviour; reproductive behaviour; fertility;
KW hormone secretion; ss.
XX

```

OS Rattus sp.

XX Key Location/Qualifiers
FH 138..1061
FT /*tag- a
FT
XX
PN W09714790-A1.
XX
PD 24-APR-1997.
XX
PF 18-OCT-1996; 96WO-US16637.
XX
PR 19-OCT-1995; 95US-0005698.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Axel R, Dulac C;
XX
DR WPI; 1997-245107/22.
DR P-PSDB; AAM19108.
XX

Nucleic acid molecule encoding vertebrate pheromone receptor -
useful to identify modulators for control of reproductive and social
behaviour, fertility and hormone secretion

XX Claim 1; Fig 13; 123pp; English.

XX cDNA clones (AAT69546-50) respectively code for rat pheromone
receptors VN3, VN4, VN5, VN6 and VN7 (AAM19104-08), members of a
novel family of presumed 7-transmembrane domain receptors that are
evolutionary independent of the odorant receptors of the main
olfactory epithelium. These clones, and a clone for VN2 (sequence
not provided), were isolated from rat vomeronasal organ cDNA
libraries by PCR and hybridisation. A differential cloning method
has been used to isolate VN7 cDNA (AAT69545). VN1-VN7 (not VN2) clones
have been deposited in pBluescript as ATCC 97294-97299. A human
homologue, HG25 (AAT69551), has also been isolated. VN polypeptides
can be expressed in host-vector systems for use in identifying
modulators for control of maternal, reproductive and social
behaviour, to increase fertility, control hormone secretion and to
regulate food uptake in humans and animals.

XX Sequence 1265 BP; 325 A; 287 C; 226 G; 426 T; 1 other;
SQ

Query Match 18.3%; Score 204.4; DB 18; Length 1265;
Best Local Similarity 55.6%; Pred. NO. 4.1e-47;
Matches 497; Conservative 0; Mismatches 386; Indels 11; Gaps 5;

QY 122 atgataaacttctcgtgttttattaccatacaaaattctttttatccacaagcagcttt 181
Db 138 atgatgaacctgtctctatggtccagatgacaaacatgatctcatcagggactgtt 197
QY 182 ggaatttcagcaaacaccatccttctttccacatcttcaccttctgttttcagtcac 241
Db 198 cgcaccttccctaacagatcctgttttttggccacctgcagctgttctttgaagaagac 257
QY 242 aggtctaagtcattgacatgataaattagtcacctgtctctatccacatactgtgctc 301
Db 258 aggtctaagccattgatctgtgcatgttcttcttattccttaacccaactaaagtcttt 317
QY 302 ttactcaggcaaatattggtgtctcttagacttttggttccacagaatactcaggatgat 361
Db 318 gtaactatgggactcatag---ctgcagacatgtttatggtcagggaataggatatt 374
QY 362 cttaggtataaggtcattgtctttttaaacaaaggttgatggggcctctccatctgacc 421
Db 375 accacatgcaggctcccttatttatttccacagaccttttgggggttttcaacctttgct 434
QY 422 cctgtcctctcagtggtgtccagggccatca---tcagccccagcatcttctccttgga 478
Db 435 gctgtctactgcatactcctttgagaccttccactcactcagtcagtagaagctcctgtttaaca 494

Db 366 tgtcttcagtagtagacataatattcttactgtggtgttcttccattgtgaccagacca 425
Qy 601 gtctctgtttgtcatacagcagctgttctctttttgtccatgagctacaccacagagagac 660
Db 426 atctactaaggtccgcaaatactgtctcaogtcttcccatgaaatccatcatgtggggag 485
Qy 661 tgttttcacactaatgactttgaggtagtgcacatttatagggttcatgggtctctcaa 720
Db 486 tgttttcc-----ttgtaggattacgctgctctcaa 516
Qy 721 gaggtcacatggtgattattttatcacagacaataagaggtctatctcagtgcttcacgca 780
Db 517 gtgcatacatgatgatctttttgtcca-agcatcagaagtgcacccagcatcttcacagt 575
Qy 781 gcaagcgtgtccc--gagttcacaacagtgaaagagccctccacaggtctatcttactgctg 838
Db 576 accagcgttccccaagatctgcagagaaaaggggttaccagagatcctctgccaactg 635
Qy 839 gtgagtt---tgtcttcacatactgggtgagactttacgttctctcatatttcagaggtgtg 895
Db 636 gtgaattgtttgtcagttgt 695
Qy 896 acatggataaatgattctgtctgt 955
Db 696 ttatggagctataaccagtcactctgagcatctagaacctgtgtgtgtgtgtgtgtgtgtgt 755
Qy 956 gcaattagctcttgatgctaattatgtcgtgataaccaataatattcagagactctgcaaatg 1015
Db 756 actctgtctctgt 815
Qy 1016 ttatggttttaaat 1028
Db 816 atggaataaagt 828

RESULT 13
AAD24968
ID AAD24968 standard; cDNA; 1015 BP.
XX AC AAD24968;
XX DT 12-MAR-2002 (first entry)
XX DE Human G-protein coupled receptor-13 (GCREC-13) cDNA.
XX KW Human; G-protein coupled receptor-13; GCREC-13; therapy; cancer; stroke;
cell proliferative disorder; neurological; epilepsy; Parkinson's disease;
Alzheimer's disease; inflammation; thyroiditis; haemolytic anaemia; AIDS;
Acquired Immune Deficiency Syndrome; dementia; nontropic; cholelithiasis;
multiple sclerosis; atherosclerosis; angina pectoris; gastroenteritis;
diabetes; ulcer; viral infection; immunosuppressive; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
CDS 248..1015
FT /*tag= a
FT /product= "Human GCREC-13 protein"
FT
XX W0200198351-A2.
XX PN 27-DEC-2001.
XX PR 15-JUN-2001; 2001WO-US19275.
XX PR 16-JUN-2000; 2000US-212483P.
XX PR 22-JUN-2000; 2000US-213954P.
XX PR 29-JUN-2000; 2000US-215209P.
XX PR 07-JUL-2000; 2000US-216595P.
XX PR 14-JUL-2000; 2000US-218936P.
XX PR 19-JUL-2000; 2000US-219154P.
XX PR 21-JUL-2000; 2000US-220141P.

PA (INCY-) INCYTE GENOMICS INC.
XX Lal P, Baughn MR, Hafalia AJA, Nguyen DB, Gandhi AR, Kallick DA;
PI Griffin JA, Yue H, Khan FA, Patterson C, Lu DAM, Tribouley CM;
PI Lu Y, Walia NK, Graul R, Yao MG, Yang J, Ramkumar J, Au-Young J;
PI Elliott VS, Hernandez R, Walsh RT, Borowsky ML, Thornton M, He A;
XX WPI; 2002-075627/10.
DR P-PSDB; AAE15643.
XX Isolated human G-protein coupled receptor polypeptides and the use of
PT these sequences in the diagnosis, treatment and prevention of diseases
PT and in the assessment of exogenous compounds on the expression of the
PT receptors
XX Claim 11; Page 139; 143pp; English.
XX The invention relates to isolated human G-protein coupled receptor
CC (GCREC) polypeptides and their biologically active fragments. GCREC and
CC protein is useful in treating a disease or condition associated with an
CC increase or decrease in expression of functional GCREC. The GCREC's are
CC useful in the diagnosis, treatment and prevention of cell proliferative
CC disorders (cancer, leukaemia, melanoma); neurological disorders (stroke,
CC epilepsy, Parkinson's disease, dementia; Alzheimer's disease); autoimmune
CC inflammatory disorder (thyroiditis, haemolytic anaemia, AIDS, multiple
CC sclerosis); cardiovascular disorder (atherosclerosis, angina pectoris),
CC gastrointestinal disorder (ulcer, cholelithiasis, gastroenteritis),
CC metabolic disorders (diabetes); viral infections (herpes virus) and in
CC the assessment of the effects of exogenous compounds on the expression
CC of the nucleic acid and amino acid sequences. The present sequence is
CC human GCREC-13 cDNA.
XX SQ Sequence 1015 BP; 261 A; 226 C; 182 G; 346 T; 0 other;

Query Match 16.3%; Score 181.6; DB 24; Length 1015;
Best Local Similarity 58.9%; Pred. No. 9.2e-41;
Matches 468; Conservative 0; Mismatches 299; Indels 27; Gaps 8;
Qy 130 acttcctgggtttattaccatacaaaatcttctttttatccacaagcagcttgggaatttc 189
Db 222 actttgacattgtgtcccatgaaaaatgccttctcattcttaagctggtattggcatctc 281
Qy 190 agcaaacacacatcctcttcttttccacatcttcacctttg-----ttttcagtcaca 242
Db 282 agcaaacacactttctctcttcttctgtctctctctctctctctctctctctctctca 341
Qy 243 ggttaagtccattgacatgataattagtcacgtctctctctctctctctctctctctct 302
Db 342 ggcagagcgcacttaccatctgctgctgctgctgctgctgctgctgctgctgctc 401
Qy 303 tcaatcaggcaaatattggtctctgactctcttgggttcacagaataactcagagatgac 362
Db 402 tcaccatgg---tgttctgtctccacagctctttgaaatcactgaatttccagaatgact 458
Qy 363 ttaggtataaggtcatgtcttttttaaacaggtgaggggctctccatctgcaccc 422
Db 459 tcaaatatagggtcatcttctctcctgaggagggtgagcaggagcctctccattgtacca 518
Qy 423 cctgctcctcgtggtgtctccag---gccatcatcagcccccagcatcttctccttggcaa 479
Db 519 cctgctcctcgtggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctga 578
Qy 480 agtcaaacacatcttctgcaagtccacatctcttaggattcttctctctctctctctctca 539
Db 579 ggtttaaatggaaatccacaatttttacccttccattgttctctcatggctc-----tctcag 633
Qy 540 acatgttcattggtgtaattcttctgctgtacactgcggctacccccagtgaaacggggcc 599
Db 634 ttttctgttagtagcctgactcttttacactgtggtcttctcccaatgtgacccag-- 691
Qy 600 agtctctgttctgcatcacagcactgttctcttctgtccatgagctacacccacaggaga 659

Db 692 atcaattgcatgtcagtaataactgttcaactttctcccaataaaactcataatcagagga 751
Qy 660 ctgttttccacactaagactttgagggatgtcacctttataggtttcaggtctctctca 719
Db 752 ctgttttccactgtcattattcagagatgtttttttaacaaaataatgctgttctca 811
Qy 720 agaggtacatggtgattttttatcacagacaataaagaggtatctcagtgcccttcacgc 779
Db 812 agtctacatgatgactctcattcaggaaactacagag---atcctggtaacctcag 868
Qy 780 agccagctgtcccgaggtctcacagtgaaaagag-cctcccaaggtatcttactgctg 838
Db 869 cccagcctctacctaaggtatcttgcagaggccaagagccatcagcacatcctgctccg 928
Qy 839 gtgagttt---tgtcttcacatactggtggaactttacgttctctcattttcagggaggtg 895
Db 929 gtgagtttctcgggtggcatgtacaagatggagcttcacatctcaacacctcctcaactg 988
Qy 896 acatggataaataga 909
Db 989 ccatgggcatatga 1002

RESULT 14
AAH74904
ID AAH74904 standard; DNA; 628 BP.
AC AAH74904;
XX
XX
DT 29-OCT-2001 (first entry)
XX
XX
DE Probe used to isolate human VNO cDNA sequences.
KW Human; vomeronasal organ; VNO; pheromone receptor; ion channel;
KW pheromone; vomeropherin; probe; ss.
XX
OS Homo sapiens.
XX
PN WO200161046-A2.
XX
XX
PD 23-AUG-2001.
XX
XX
PF 15-FEB-2001; 2001WO-US05178.
XX
XX
PR 17-FEB-2000; 2000US-0183128.
XX
XX
PA (PHER-) PHERIN PHARM INC.
XX
XX
PI Herman RC, Berliner D;
XX
XX
DR WPI; 2001-522609/57.
XX
XX
PT Novel vomeronasal organ cDNA library useful for identification and
XX isolation of genes expressed in vomeronasal organs, specifically genes
XX for pheromone receptors and ion channels
XX
XX
PS Example 9; Page 54-55; 59pp; English.
XX
XX
CC The present sequence represents a probe, used to isolate human
CC vomeronasal organ (VNO) cDNA sequences. The cDNA library was
CC constructed from female tissue. Sequences isolated from the VNO cDNA
CC library are useful as a unique resource for the identification and
CC isolation of genes expressed in VNO, specifically genes for pheromone
CC receptors, ion channels and prospective reagents for high throughput
CC assays. They are useful as an excellent source to search for novel
CC genes, gene fragments or other nucleotide sequences encoding proteins
CC that are implicated in detection of pheromones or other vomeropherins
CC in the human VNO. The cDNAs are useful as hybridization probes for
CC determining the presence or concentration of an oligo- or polynucleotide,
CC e.g. DNA, of interest.
XX
XX
SQ Sequence 628 BP; 138 A; 178 C; 123 G; 189 T; 0 other;

Query Match 14.6%; Score 163.2; DB 22; Length 628;
Best Local Similarity 60.1%; Pred No. 1.1e-35;
Matches 380; Conservative 0; Mismatches 238; Indels 14; Gaps 6;
Qy 258 acatgataatagtcacactgtctctcatccacatactgctgctcttccactcaggaatat 317
Db 1 acctgacatcagctctcttggccctccaccttggatgc---taacagtcacgtgat 57
Qy 318 tgggtccttagactcttcttgggttcacagaaatactcaggtatgactttaggtataaggtca 377
Db 58 tcagagctgtgtatttttgcattctcagaatgtgtggaatgacatcaaatgcaatcccc 117
Qy 378 ttgtcttttaaaacaaggtgatgaggggctctccatctgcacccctgcctcctcctgagtg 437
Db 118 ttgccacttacacagacttttggagggcctctctcttctgtactctgtctgtgagta 177
Qy 438 tgcctcaggccatca---tcagcccccgcactctctctctcttggcaagctcaaacatcctt 494
Db 178 tcttccaggccatcacctcttagccccaagactctctgttttagcaagtccaataataat 237
Qy 495 ctgcaagtcacatcttaggattcttctcttctctcatggtctcacaatgttctcattgtg 554
Db 238 ccacacagcacagcctgttctcctctgtctgtctgtggcctctctacatgctcgtgta 297
Qy 555 taattctctgtgtac--actgcgggtcaccccccagtgaaacggggccagctcttctgtttg 612
Db 298 ctcaactactcttcacacactgtgtgactacaaactctcttcaacgagctcactatgtg 357
Qy 613 tcatacagcactgtctcttcttggccatgagctacacccacagagagactgttttccacac 672
Db 358 tcactgaatcctgattattttacccatggtattacacaccagggatttttttccatat 417
Qy 673 taatgactttgagggatgtcacctttataggtttcaggtctcctcctcaagagctacatgg 732
Db 418 tggggatatttcgggtgtgtctctcataggtctcctcctcctcctcaggggacatgg 477
Qy 733 tgattattttacagacaataagaggtctatctcagctgccttctcagcagcagcctgtcc 792
Db 478 tggccctctgtgcagaca-caggaaacaggccacagcatcttccacaggaccagccttct 536
Qy 793 cc--gagctcaccagtgaaaagagcctccagcgtatcttactgtctgtgag---tttt 847
Db 537 ccaaaagcatcccccagagcaaggccacacaggaccatcctgttgcctcaggtctttt 596
Qy 848 gtcttcacatctggtgagctttacgtttctc 879
Db 597 gtgtgagtactgtgtgagctgcaccatc 628

RESULT 15
AAH74901
ID AAH74901 standard; DNA; 632 BP.
AC AAH74901;
XX
XX
DT 29-OCT-2001 (first entry)
XX
XX
DE Probe used to isolate human VNO cDNA sequences.
KW Human; vomeronasal organ; VNO; pheromone receptor; ion channel;
KW pheromone; vomeropherin; probe; ss.
XX
OS Homo sapiens.
XX
PN WO200161046-A2.
XX
XX
PD 23-AUG-2001.
XX
XX
PF 15-FEB-2001; 2001WO-US05178.
XX
XX
PR 17-FEB-2000; 2000US-0183128.
XX

PA (PHER-) PHERIN PHARM INC.

PI Herman RC, Berliner D:

XX
DR
WPI; 2001-522609/57.

Novel vomeronasal organ cDNA library useful for identification and isolation of genes expressed in vomeronasal organs, specifically genes for pheromone receptors and ion channels -

PS Example 9; Page 52-53; 59pp; English.

The present sequence represents a probe, used to isolate human vomeronasal organ (VNO) cDNA sequences. The cDNA library was constructed from female tissue. Sequences isolated from the VNO cDNA library are useful as a unique resource for the identification and isolation of genes expressed in VNO, specifically genes for pheromone receptors, ion channels and prospective reagents for high throughput assays. They are useful as an excellent source to search for novel genes, gene fragments or other nucleotide sequences encoding proteins that are implicated in detection of pheromones or other vomeropherins in the human VNO. The cDNAs are useful as hybridization probes for determining the presence or concentration of an oligo- or polynucleotide e.g. DNA, of interest.

Sequence 632 BP; 150 A; 177 C; 109 G; 196 T; 0 other;

Query Match	14.28;	Score 158.4;	DB 22;	Length 632;
-------------	--------	--------------	--------	-------------

Best Local Similarity 59.78; Pred. No. 2.4e-34; Score 133.4; SD 14.20; Accuracy 100%;

2000 Seed Library 33.7%, Acc. No. 2.46 34,
 Matches 358; Conservative 0; Mismatches 231; Indels 11; Gaps 5;

Qy 261 tgataattagtcacctgtctctcatccacatactgctgctcttcactcaggaatttg 320

Db 1 tggccattgggtctctttgtccctaatccactactgctactgctatgqgggc---attca 57

Qy 321 tgtccttagacttcttttggttcacagaataactcaqqatgaccttagqataaqqtcattg 380

58 tagccatagacatcttttatttcttggaggggcatgggatcacatcatatgtatgaaatcccttg 117

Qy 381 tctttttaacaaggatgattgaaggagccctctccatctgcacccccctgccctcctgaagtatgc 440

47
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

441 tccaggccatca--tcagcccccagcatctctccttgcacaaagctcaaacatccttctc 497

Accession	Gene	Accession	Gene
U00096	16S rRNA	U00096	16S rRNA
U00097	23S rRNA	U00097	23S rRNA
U00098	5S rRNA	U00098	5S rRNA
U00099	16S rRNA	U00099	16S rRNA
U00100	23S rRNA	U00100	23S rRNA
U00101	5S rRNA	U00101	5S rRNA
U00102	16S rRNA	U00102	16S rRNA
U00103	23S rRNA	U00103	23S rRNA
U00104	5S rRNA	U00104	5S rRNA
U00105	16S rRNA	U00105	16S rRNA
U00106	23S rRNA	U00106	23S rRNA
U00107	5S rRNA	U00107	5S rRNA
U00108	16S rRNA	U00108	16S rRNA
U00109	23S rRNA	U00109	23S rRNA
U00110	5S rRNA	U00110	5S rRNA
U00111	16S rRNA	U00111	16S rRNA
U00112	23S rRNA	U00112	23S rRNA
U00113	5S rRNA	U00113	5S rRNA
U00114	16S rRNA	U00114	16S rRNA
U00115	23S rRNA	U00115	23S rRNA
U00116	5S rRNA	U00116	5S rRNA
U00117	16S rRNA	U00117	16S rRNA
U00118	23S rRNA	U00118	23S rRNA
U00119	5S rRNA	U00119	5S rRNA
U00120	16S rRNA	U00120	16S rRNA
U00121	23S rRNA	U00121	23S rRNA
U00122	5S rRNA	U00122	5S rRNA
U00123	16S rRNA	U00123	16S rRNA
U00124	23S rRNA	U00124	23S rRNA
U00125	5S rRNA	U00125	5S rRNA
U00126	16S rRNA	U00126	16S rRNA
U00127	23S rRNA	U00127	23S rRNA
U00128	5S rRNA	U00128	5S rRNA
U00129	16S rRNA	U00129	16S rRNA
U00130	23S rRNA	U00130	23S rRNA
U00131	5S rRNA	U00131	5S rRNA
U00132	16S rRNA	U00132	16S rRNA
U00133	23S rRNA	U00133	23S rRNA
U00134	5S rRNA	U00134	5S rRNA
U00135	16S rRNA	U00135	16S rRNA
U00136	23S rRNA	U00136	23S rRNA
U00137	5S rRNA	U00137	5S rRNA
U00138	16S rRNA	U00138	16S rRNA
U00139	23S rRNA	U00139	23S rRNA
U00140	5S rRNA	U00140	5S rRNA
U00141	16S rRNA	U00141	16S rRNA
U00142	23S rRNA	U00142	23S rRNA
U00143	5S rRNA	U00143	5S rRNA
U00144	16S rRNA	U00144	16S rRNA
U00145	23S rRNA	U00145	23S rRNA
U00146	5S rRNA	U00146	5S rRNA
U00147	16S rRNA	U00147	16S rRNA
U00148	23S rRNA	U00148	23S rRNA
U00149	5S rRNA	U00149	5S rRNA
U00150	16S rRNA	U00150	16S rRNA
U00151	23S rRNA	U00151	23S rRNA
U00152	5S rRNA	U00152	5S rRNA
U00153	16S rRNA	U00153	16S rRNA
U00154	23S rRNA	U00154	23S rRNA
U00155	5S rRNA	U00155	5S rRNA
U00156	16S rRNA	U00156	16S rRNA
U00157	23S rRNA	U00157	23S rRNA
U00158	5S rRNA	U00158	5S rRNA
U00159	16S rRNA	U00159	16S rRNA
U00160	23S rRNA	U00160	23S rRNA
U00161	5S rRNA	U00161	5S rRNA
U00162	16S rRNA	U00162	16S rRNA
U00163	23S rRNA	U00163	23S rRNA
U00164	5S rRNA	U00164	5S rRNA
U00165	16S rRNA	U00165	16S rRNA
U00166	23S rRNA	U00166	23S rRNA
U00167	5S rRNA	U00167	5S rRNA
U00168	16S rRNA	U00168	16S rRNA
U00169	23S rRNA	U00169	23S rRNA
U00170	5S rRNA	U00170	5S rRNA
U00171	16S rRNA	U00171	16S rRNA
U00172	23S rRNA	U00172	23S rRNA
U00173</			

[illegible]

QY 496 caagtcacatctcaggattcccccctcccccaaacatggtccatgggtgttaa 35
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DH 238 cccatacacgttgcctgtttgccattatgttcgaagcgactcatagaccacagt 207

[illegible]

DH	accctttagtatccatcgaacgcccccccgccttcagcttcgtttgca	357
QY	cctccctctgtacacct--gccggcacccccccaggaaacggggcccgagtcttcgtttgcta	619

[illegible][illegible]

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523</
--	---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-------

6/6 **cgactttgagggatgcacrtttatagggttcattggtcctctcaagaggctacatgggtga** 735

[illegible]

Qy 736 ttattttatacagacaaataagaggctatctcagtcgcttcacgcagccagcctgtcccc- 794

Qy 795 -gagtctaccagtgaaagagcctccaggctatcttactgctggtgagtttgccttc 853

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 04:00:29 ; Search time 4107.22 Seconds
(without alignments)
5675.898 Million cell updates/sec

Title: US-09-728-309-2
Perfect score: 1114
Sequence: 1 attccagatcatagatgt.....gagaggttaatccattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_lm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_un.*
28: em_vi.*
29: em_hum.*
30: em_htg.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----	-----	-----	-----	-----	-----

1	1098	98.6	206832	2	AL390728	AL390728 Homo sapi
2	1094.8	98.3	156798	30	AL359674	AL359674 Human DNA
3	1094.8	98.3	217087	2	AC104335	AC104335 Homo sapi
4	542.6	48.7	206832	2	AL390728	AL390728 Homo sapi
5	375.6	33.7	217087	2	AC104335	AC104335 Homo sapi
6	313.2	28.1	73661	9	AC074050	AC074050 Homo sapi
7	313.2	28.1	152748	9	AC106730	AC106730 Homo sapi
8	304.6	27.3	65128	9	AC073188	AC073188 Homo sapi
9	304.6	27.3	139384	9	AC006455	AC006455 Homo sapi
10	303.6	27.3	989	4	AB064662	AB064662 Capra hlr
11	281.2	25.2	151941	9	AC108131	AC108131 Homo sapi
12	281.2	25.2	161609	2	AC011823	AC011823 Homo sapi
13	280.2	25.2	95097	9	AC022202	AC022202 Homo sapi
14	279.6	25.1	65128	9	AC073188	AC073188 Homo sapi
15	279.6	25.1	149495	30	AC022860	AC022860 Homo sapi
16	275.4	24.7	194142	9	AC092634	AC092634 Homo sapi
17	268.2	24.1	68299	9	AC091685	AC091685 Homo sapi
18	268.2	24.1	152964	2	AC011821	AC011821 Homo sapi
19	267.2	24.0	909	10	AY065469	AY065469 Mus muscu
20	265.6	23.8	909	10	AY065471	AY065471 Mus muscu
21	262	23.5	178105	9	AC073057	AC073057 Homo sapi
22	255.6	22.9	891	10	AY065468	AY065468 Mus muscu
23	251.6	22.6	903	10	AF291501	AF291501 Mus muscu
24	250.8	22.5	909	10	AY065466	AY065466 Mus muscu
25	249.8	22.4	900	10	AF291499	AF291499 Mus muscu
26	248.4	22.3	161571	9	AC069462	AC069462 Homo sapi
27	248.4	22.3	162339	2	AC024151	AC024151 Homo sapi
28	248.4	22.3	166991	9	AC090884	AC090884 Homo sapi
29	247.6	22.2	918	10	AY065467	AY065467 Mus muscu
30	246.2	22.1	1020	10	AY065473	AY065473 Mus muscu
31	244.8	22.0	188389	2	AC024199	AC024199 Homo sapi
32	244.8	22.0	195905	2	AC092447	AC092447 Homo sapi
33	244.4	21.9	894	10	AY065477	AY065477 Mus muscu
34	244.4	21.9	918	10	AY065470	AY065470 Mus muscu
35	242.6	21.8	152964	2	AC011821	AC011821 Homo sapi
36	241.2	21.7	912	10	AY065489	AY065489 Mus muscu
37	240.6	21.6	891	10	AY065474	AY065474 Mus muscu
38	240.4	21.6	202333	2	AC080144	AC080144 Mus muscu
39	239.6	21.5	1738	9	AF253312	AF253312 Homo sapi
40	239.4	21.5	957	10	AB062896	AB062896 Mus muscu
41	238.8	21.4	3403	10	MMPEREC2	Y12724 Mus musculu
42	238	21.4	897	10	AY065487	AY065487 Mus muscu
43	237.6	21.3	912	10	AY065488	AY065488 Mus muscu
44	237.4	21.3	900	10	AY065486	AY065486 Mus muscu
45	236.2	21.2	223275	2	AC098739	AC098739 Mus muscu

ALIGNMENTS

RESULT 1

AL390728
LOCUS AL390728 206832 bp DNA linear HTG 06-FEB-2002
DEFINITION Homo sapiens chromosome 1 clone RP11-488L18, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.

ACCESSION AL390728
VERSION AL390728.22 GI:18614040
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Ashwell, R.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 7, 2002 this sequence version replaced gi:18491329.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: bA488LJ8
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Et-amersham; 9% of reads Chemistry:
 Dye-terminator Big Dye; 90% of reads
 Consensus quality: 202499 bases at least Q40
 Consensus quality: 203211 bases at least Q30
 Consensus quality: 203559 bases at least Q20
 Insert size: 206732; sum-of-contigs
 Insert size: 205638; 4.3% error; agarose-fp
 Quality coverage: 9.91x in Q20 bases; sum-of-contigs Quality
 coverage: 10.88x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 204689: contig of 204689 bp in length
 * 204690 204789: gap of 100 bp
 * 204790 206832: contig of 2043 bp in length.
 FEATURES
 source
 1. .206832
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone_lib="RP11-488L18"
 /clone_lib="RP11-11.2"
 misc_feature
 1. .204689
 /note="assembly_fragment:03301"
 misc_feature
 204790..206832
 /note="assembly_fragment:03801"
 BASE COUNT 58135 a 44577 c 44388 g 59631 t 101 others
 ORIGIN
 Query Match 98.6%; Score 1098; DB 2; Length 206832;
 Best Local Similarity 99.7%; Pred. No. 3e-273;
 Matches 1111; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
 QY 1 attccagatcatagagatgttgaaattggtattattgagaacatgcagagaaattatgct 60
 Db 144680 ATTCACATCATAGAGATGTTGAATTTGGTTATTATTGAGAACATGCGACAAATATATGCT 144739
 QY 61 attctattagatctctgtttttctccacagatatccttctgttttaatttctctctaa 120
 Db 144740 ATTCATATTAGATCTCTGTCTTTCTCCACAGATATCCTTTGCTTTAATTTCTCTTAA 144799
 QY 121 gatgatacaacttctctgtttttattacacatacaaaattcttctttatccacaagccagctt 180
 Db 144800 GATGATCAAACTTCTCTGTTTATTACCATACAAATCTCTTTTATCCACAAGCCAGCTT 144859
 QY 181 tggatttcagcaaacaccatctcttctttccacatcttccacttttttttttttttttttttt 240
 Db 144860 TGGAAATTCAGCAAAACACCATCTCTCTCTTTCCACATCTTACACCTTTGTTTTCAGTCA 144919
 QY 241 caggtctaaagtccattgacatgataattagtcacctgtctctatccacacatactgtgct 300
 Db 144920 CAGGTCTAAAGTCCATTGACATGATAATTAGTCACCTGTCTCTCATCCACATCTGTCTGCT 144979
 QY 301 cttcactcaggcaaatatggtgtctcttagacttttttttttttttttttttttttttttttt 360
 Db 144980 CTTCACTCAGGCAATATTTGGTCTCTTAGACTTTCTTTTCCACATCTTACACCTTTGTTTTCAGTCA 145039
 QY 361 tcttagtataaagtcattgtcttttttaacaaggtgataggggctctccatctcac 420
 Db 145040 TCTTAGGTATAAGGTCATTGTCTTTTAAACAAAGGTGATAGGGGCTCTCCATCTCCAC 145099

QY 421 cccctgctcctcagtgctccaggccatcatcagcccccagcatcttctctctggcaaa 480
 Db 145100 CCCCTGCTCTCTGAGTGTGTCTCCAGGCCATCATCAGCCCCAGCATCTCTCTCTGGCAAA 145159
 QY 481 gctcaaacatcctcttctgcaagtcacatcttaggattcttctctctctctctctctcaaa 540
 Db 145160 GCTCAAAACATCTTCTGCAAGTCACATCTTAGGATTCTTCTCTCTCTCTCTCTCTCAAA 145219
 QY 541 catgttcatgggtgtaattcttctgctgtacactgctgctacccccccagtggaaggggcca 600
 Db 145220 CATGTTCAATTTGGTGAATCTTCTGTACACTGCGCTACCCCCAGTGAACGGGGCCA 145279
 QY 601 gtctctgttcttcatacacagcactgttctcttttggccatgagctacacccacagagac 660
 Db 145280 GTCTTCTGTTTGTTCATACAGCAGCTGTTCTTTTGGCCATGAGCTACACCCACAGGAGAC 145339
 QY 661 tgttttcacactaatgacttttgagggatgctacacctttatagggttcatggtctctcaaa 720
 Db 145340 TGTTTTTCACACTAATGACTTTGAGGGATGTCACCTTTATAGGGTTTCATGCTCTCTCAAA 145399
 QY 721 gaggtcacatgggtgattattttatcacagacaataagaggctatctcagtcgcttcacaca 780
 Db 145400 GAGGCTACATGGTGATTATTTTATACAGACAATTAAGAGGCTATCTCAGTGCCCTTCACACA 145459
 QY 781 gccagcctgtcccgagctctcaccagtgaaaagagccctccaggtctatcttactctgctgt 840
 Db 145460 GCCAGCCTGTCCCGAGTCTCACCCAGTGAAGAGCCTCCAGGCTATCTTACTGTCTGT 145519
 QY 841 gagtttcttccacatactggttgagcttcttctctctctctctctctctctctctctctct 900
 Db 145520 GAGTTTGTCTTTCACATCTGCTGAGTGGTGGACTTTTACGTTCTCTCAATTTTTCAGGAGGTGTGACATG 145579
 QY 901 gataaatgattctctgctagtgctgctccaggttattgtgccaatagctatgctccgcaat 960
 Db 145580 GATAAATGATTCTCTGCTAGTGGCTCCAGGTTATTGTGGCCAATAGCTATGCGCGCAAT 145639
 QY 961 tagctctttgatgctaatattatgctgataaccaaattattcaagactctgcaaatgtttatg 1020
 Db 145640 TAGTCTCTTGTATGCTAATTTATGCTGATAACCAAAATATTCAAGACTCTGCAAAATGTTATG 145699
 QY 1021 gtttaaatatttctctcctccaaagctcatgttgaaatttaataatgccaatgtggcagtac 1080
 Db 145700 GTTTAAATATTGTCTCTCTCCAAAGCTCATGTTGAAATTTTAATGCGCAATGTGCGAGTAC 145759
 QY 1081 taagaaatgatgatgagaggttaattccattcatg 1114
 Db 145760 TAAGAAG---TGATGAGAGGTTAATCCATTTCATG 145790
 RESULT 2
 AL359674/C
 ID AL359674 standard; DNA; HTG; 166798 BP.
 XX AC AL359674;
 XX SV AL359674.4
 XX DT 20-JUN-2000 (Rel. 64, Created)
 XX DT 20-JAN-2001 (Rel. 66, Last updated, Version 5)
 XX DE Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone RP11-433K2
 XX KW HTG; HTGS_DRAFT; HTGS_PHASE1.
 XX OS Homo sapiens (human)
 XX OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
 XX OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 XX RN [1]
 XX RA Plumb B.;
 RT Submitted (20-JAN-2001) to the EMBL/GenBank/DBJ databases.

RL Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 RL humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 XX

CC ----- Genome Center

CC Center: Sanger Centre

CC Center code: SC

CC Web site: <http://www.sanger.ac.uk>

CC Contact: humquery@sanger.ac.uk

CC ----- Project Information

CC Center project name: BA433K2

CC ----- Summary Statistics

CC Assembly program: XGAP4; version 4.5

CC Sequencing vector: plasmid: L08752; 100% of reads

CC Chemistry: Dye-terminator Big Dye; 100% of reads

CC Consensus quality: 153591 bases at least Q40

CC Consensus quality: 158804 bases at least Q30

CC Consensus quality: 161513 bases at least Q20

CC Insert size: 163998; sum-of-contigs

CC Insert size: 213184; 4.3% error; agarose-fp

CC Quality coverage: 3.38x in Q20 bases; sum-of-contigs

CC Quality coverage: 2.98x in Q20 bases; agarose-fp

CC -----
 CC * NOTE: This is a 'working draft' sequence. It currently
 CC * consists of 29 contigs. The true order of the pieces is
 CC * not known and their order in this sequence record is
 CC * arbitrary. Where the contigs adjacent to the vector can
 CC * be identified, they are labelled with 'clone_end' in the
 CC * feature table. Some order and orientation information
 CC * can tentatively be deduced from paired sequencing reads
 CC * which have been identified to span the gap between two
 CC * contigs. These are labelled as part of the same
 CC * 'fragment_chain', and the order and relative orientation
 CC * of the pieces within a fragment_chain is reflected in
 CC * this file. Gaps between contigs are represented as
 CC * runs of N, but the exact sizes of the gaps are unknown.
 CC * This record will be updated with the finished sequence as
 CC * soon as it is available and the accession number will be
 CC * preserved.

CC 1 2892 contig of 2892 bp in length; fragment_chain 1
 CC 2 6216 contig of 3224 bp in length; fragment_chain 1
 CC 3 6317 11678 contig of 5362 bp in length; fragment_chain 1
 CC 4 11779 23525 contig of 11747 bp in length; fragment_chain 1
 CC 5 23626 28320 contig of 4695 bp in length; fragment_chain 1
 CC 6 28421 40356 contig of 11936 bp in length; fragment_chain 1
 CC 7 40457 45185 contig of 4729 bp in length; fragment_chain 2
 CC 8 45286 51072 contig of 5787 bp in length; fragment_chain 2
 CC 9 51173 58120 contig of 6948 bp in length; fragment_chain 2
 CC 10 58221 72019 contig of 3799 bp in length; fragment_chain 3
 CC 11 62120 74687 contig of 12568 bp in length; fragment_chain 3
 CC 12 74788 79347 contig of 4560 bp in length; fragment_chain 4
 CC 13 79448 84093 contig of 4846 bp in length; fragment_chain 4
 CC 14 84194 86233 contig of 2040 bp in length; fragment_chain 5
 CC 15 86334 89925 contig of 3592 bp in length; fragment_chain 5
 CC 16 90026 95379 contig of 5354 bp in length; fragment_chain 6
 CC 17 95480 105307 contig of 9828 bp in length; fragment_chain 6
 CC 18 105408 123245 contig of 17838 bp in length; fragment_chain 7
 CC 19 123346 125459 contig of 2114 bp in length; fragment_chain 7
 CC 20 134133 136801 contig of 8473 bp in length; fragment_chain 8
 CC 21 136902 142924 contig of 2669 bp in length; fragment_chain 8
 CC 22 143025 145813 contig of 6023 bp in length; fragment_chain 9
 CC 23 145914 148777 contig of 2789 bp in length; fragment_chain 9
 CC 24 148878 151529 contig of 2864 bp in length
 CC 25 151630 155541 contig of 2652 bp in length
 CC 26 155642 159186 contig of 3912 bp in length
 CC 27 159287 162766 contig of 3545 bp in length
 CC 28 162867 166798 contig of 3480 bp in length
 CC 29 166798 contig of 3932 bp in length

XX Key Location/Qualifiers

FH source

FT 1. .166798

FT /chromosome="1"

FT /db_xref="taxon:9606"

FT

FT /organism="Homo sapiens"
 FT /clone="Rp11-433K2"
 FT /clone_lib="RPC1-11.2"
 FT 1. .2892
 FT /note="assembly_fragment:00889"
 FT /note="fragment_chain:1"
 FT 2993. .6216
 FT /note="assembly_fragment:01566"
 FT /note="fragment_chain:1"
 FT 6317. .11678
 FT /note="assembly_fragment:00917"
 FT /note="fragment_chain:1"
 FT 11779. .23525
 FT /note="assembly_fragment:00101"
 FT /note="fragment_chain:1"
 FT 23626. .28320
 FT /note="assembly_fragment:00080"
 FT /note="fragment_chain:1"
 FT 28421. .40356
 FT /note="assembly_fragment:01357"
 FT /note="fragment_chain:1"
 FT 40457. .45185
 FT /note="assembly_fragment:01837"
 FT /note="fragment_chain:2"
 FT 45286. .51072
 FT /note="assembly_fragment:01319"
 FT /note="fragment_chain:2"
 FT 51173. .58120
 FT /note="assembly_fragment:01296"
 FT /note="fragment_chain:2"
 FT 58221. .62019
 FT /note="assembly_fragment:00257"
 FT /note="fragment_chain:3"
 FT 62120. .74687
 FT /note="assembly_fragment:01472"
 FT /note="fragment_chain:3"
 FT 74788. .79347
 FT /note="assembly_fragment:00458"
 FT /note="fragment_chain:4"
 FT 79448. .84093
 FT /note="assembly_fragment:01280"
 FT /note="fragment_chain:4"
 FT 84194. .86233
 FT /note="assembly_fragment:00502"
 FT /note="fragment_chain:5"
 FT 86334. .89925
 FT /note="assembly_fragment:00888"
 FT /note="fragment_chain:5"
 FT 90026. .95379
 FT /note="assembly_fragment:01252"
 FT /note="fragment_chain:6"
 FT 95480. .105307
 FT /note="assembly_fragment:01516"
 FT /note="fragment_chain:6"
 FT 105408. .123245
 FT /note="assembly_fragment:01328"
 FT /note="fragment_chain:7"
 FT 123346. .125459
 FT /note="assembly_fragment:00630"
 FT /note="fragment_chain:7"
 FT 125560. .134032
 FT /note="assembly_fragment:01373"
 FT /note="fragment_chain:8"
 FT 134133. .136801
 FT /note="assembly_fragment:01754"
 FT /note="fragment_chain:8"
 FT 136902. .142924
 FT /note="assembly_fragment:01731"
 FT /note="fragment_chain:9"
 FT 143025. .145813
 FT /note="assembly_fragment:01551"
 FT /note="fragment_chain:9"
 FT 145914. .148777

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 4074: contig of 4074 bp in length
 4075 4174: gap of unknown length
 4175 9708: contig of 5534 bp in length
 9709 9808: gap of unknown length
 9809 13642: contig of 3834 bp in length
 13643 13742: gap of unknown length
 13743 52846: contig of 39104 bp in length
 52847 52948: gap of unknown length
 52949 140430: contig of 87484 bp in length
 140431 140530: gap of unknown length
 140531 217087: contig of 76557 bp in length.

FEATURES

source

Location/Qualifiers
 1. .217087
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-433K2"
 /clone_lib="RPC1 human BAC library 11"
 1. .4074
 /note="assembly_name:Contig168"
 4175. .9708
 /note="assembly_name:Contig169"
 9809. .13642
 /note="assembly_name:Contig170"
 13743. .52846
 /note="assembly_name:Contig171"
 52947. .140430
 /note="assembly_name:Contig172"
 140531. .217087
 /note="assembly_name:Contig173"
 60435 a 48981 c 48285 g 58724 t 662 others

BASE COUNT
ORIGIN

Query Match 98.38; Score 1094.8; DB 2; Length 217087;
 Best Local Similarity 99.68; Pred. No. 2e-272;
 Matches 1109; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

Qy 1 attccagatcatagagattggaattgggtattattgagaaacatggcagaaattatgct 60
 Db 121154 ATTCCAGATCATAGAGATTGGAATTGGTTATTATTGAGAAACATGGCAGAAATATGCT 121095
 Qy 61 attctcattagatccttgcctttctccacagatataccttgccttttaatttcctctaa 120
 Db 121094 ATTCTCATTAGATCTTGTCTTTCTCCACAGATATCCTTTTGTCTTAAATTTCTCTTAA 121035
 Qy 121 gatcatcaaacctcctggttttattaccatacaaatcttcttttatccacagcagctt 180
 Db 121034 GATGATCAAACTTCTCTGTTTTATTACCATATAAATCTCTTTTATCCACAAGCCAGCTT 120975
 Qy 181 tggaaatttcagcaaacacacctcttcttttccacatcttcacaccttggtttcagtcga 240
 Db 120974 TGGAAATTCACAAACACCATCTTCTTCTTCCACATCTTCACTTTGTTTTCAGTCA 120915
 Qy 241 caggtcgaagtcactgacatga taattagtcacctgtctctcaccacatcactgctgct 300
 Db 120914 CAGGTCTAAGTCCCATTTGACATGATAAATTAGTCAGCTGTCTCTCATCCACATACACTGCTGCT 120855
 Qy 301 ctccactcaggcaatattgggtccttagactcttgggttcacagaaatactcagatga 360
 Db 120854 CTTCCACCAGCAATATGGTGTCTTAGACTTCTTTGGTTTCAGAAATATCTCAGGATGA 120795
 Qy 361 tcttaggtataaggtcattgttctttttaacaaggtgatggggcctctccatctgcac 420
 Db 120794 TCTTAGGTATAAGTCAATGCTCTTTTAAACAAGGTGATGAGGGCCTCTCCATCTGCAC 120735
 Qy 421 ccctcgctcctcgtgagtgctccaggccatcattcagccccagcattcttctccttggcaaa 480
 Db 120734 CCCCTGCTCCTGAGTGTGCTCCAGGCCATCATCAGCCCCCAGCATCTTCTCTTGGCAAA 120675

Qy 481 gctcaaacatctctctgcaagtccacatcttaggattcttcttcttcttcttcttcttctca 540
 Db 120674 GCTCAAAACATCTCTTCTGCAAGTCACATCTTAGGATTCTTCTCTTCTCATGGTCTCA 120615
 Qy 541 catgttcattggttaatactcttctctgtacactgagggtacccacagtgaaacggggcca 600
 Db 120614 CATGTTCAATGGTGTAAATCTTCTGCTGTACACTGCGGCTACCCCGAGTGAACGGGGCA 120555
 Qy 601 gtctctgtttgttcatacagcaactgttctcttcttcttcttcttcttcttcttcttct 660
 Db 120554 GTCTTCTGTTTGTATACAGCACTGTTCCTTTTGTGCCCATGAGCTACACCCACAGGAG 120495
 Qy 661 tgttttcacactaatgactttgagggtatgtcacctttatagggttcatgtctctctca 720
 Db 120494 TGTTTTTCACACTAATGACTTTGAGGGATGTACCTTTATAGGGTTCAATGGTCTCTCA 120435
 Qy 721 gaggctacatggtgattattttttatatacagacaataagaggctatctcagtcgcttcaca 780
 Db 120434 GAGGCTACATGGTGATTATTATTATACAGACAATAAGAGGCTATCTCAGTCGCTTCA 120375
 Qy 781 gccggcctgtcccggagctctccacagtgaaaagagcctccacagcctatcttactgctggt 840
 Db 120374 GCCAGCCTGTCCCGAGTCTCACCAGTGAAGAGAGCCCTCCAGGCTATCTTACTGCTGT 120315
 Qy 841 gagtgttcttcacatactactggtgagctttacgtttcttcttcttcttcttcttcttct 900
 Db 120314 GAGTTTGTCTTCACATACTACGGTGGACTTTACGTTCTTCATTTTCAGAGGTTGTGACATG 120255
 Qy 901 gataaatgattctctctgctagtggtgctccaggtattgttgcccaatagctatgcgcgaat 960
 Db 120254 GATAAATGATTCTCTGCTAGTGTGCTGCCAGTTATTGTGGCCAATAGCTATGCCGAAT 120195
 Qy 961 tagctcttgatgcttaatttatgtgtataccaaataatctcaagactctgtcgaatgttatg 1020
 Db 120194 TAGTCCTTTGATGCTAAATTTTATGCTGATAACCAATATTCGAAGACTCTGCAAAATGTTATG 120135
 Qy 1021 gtttaaatatttctctctccaaagctcatgttgaatttataatgcctaatgtgacgtac 1080
 Db 120134 GTTTAAATATTATTGTCTCTCTCCAAAGCTCATGTTGAAATTTTAAATGCTCAATGTGGCAGTAC 120075
 Qy 1081 taagaagtgatgatgagagggttaattccattcatg 1114
 Db 120074 TAAGAAG---TGATGAGAGGTTAATCCATTTCATG 120044

RESULT 4
AL390728/c

LOCUS AL390728 206832 bp DNA linear HTG 06-FEB-2002
 DEFINITION Homo sapiens chromosome 1 clone RP11-488L18, *** SEQUENCING IN
 PROGRESS ***, 2 unordered pieces.

ACCESSION AL390728

VERSION AL390728.22 GI:18614040

KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (sites)

AUTHORS Direct Submission

TITLE Ashwell, R.

JOURNAL Submitted (05-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Feb 7, 2002 this sequence version replaced gi:18491329.

COMMENT

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA488L18
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 9% of reads Chemistry:
 Dye-terminator Big Dye; 90% of reads
 Consensus quality: 202499 bases at least Q40
 Consensus quality: 203211 bases at least Q30
 Consensus quality: 203559 bases at least Q20
 Insert size: 206732; sum-of-contigs
 Insert size: 205638; 4.3% error; agarose-fp
 Quality coverage: 9.91x in Q20 bases; sum-of-contigs Quality
 coverage: 10.88x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 204689: contig of 204689 bp in length
 * 204690 204789: gap of 100 bp
 * 204790 206832: contig of 2043 bp in length.

FEATURES

Source
 1..206832
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone_lib="RP11-488L18"
 /clone_lib="RP11-11.2"
 misc_feature
 1..204689
 /note="assembly_fragment:03301"
 204790..206832
 /note="assembly_fragment:03801"
 BASE COUNT 58135 a 44577 c 44388 g 59631 t 101 others
 ORIGIN

Query Match 48.7%; Score 542.6; DB 2; Length 206832;
 Best Local Similarity 77.7%; Pred. No. 1.6e-129;
 Matches 814; Conservative 0; Mismatches 204; Indels 29; Gaps 12;

QY 1 attccagatcagaaga---tggtgaattggtattattgagaacatggcagaaa--- 53
 Db 126599 ATTCCAGATCACAGGATTATTGAACATAATGATACATAAACAACATGGCAGAAATGT 126540
 QY 54 ---ttatgctattctcattagatctctgtctttctccacagatattctgtcttaatt 110
 Db 126539 TCCTAATGCAATTATCATTTGATGCTTAATTTTCTCTACAGATATCTTTATTTGTTT 126480
 QY 111 ttctcttaag---atgatacaactctcctgtttttattaccatacaaatctctttatc 167
 Db 126479 AATTTTCCATGGAAATGACCAAGCTTTTCAGCTATATTGTCAATAAATAATGTTATTACC 126420
 QY 168 cacaagcagcttggaatttcagacaacaccatctcttctttccacatcttccacct 227
 Db 126419 CTCAAGTCAGCTTGGAAATCTCAGCAACACCTTCTCTCTTTTCCACATCTCACTT 126360
 QY 228 ttgttttcagtcaggtctaaagtcattgacatgataattagtcacctgtctctatcc 287
 Db 126359 TTGCTTATCTACAGGCTTAAGCCCAATGACATGACCAATTAGTCACTTGCCCTAATCC 126300
 QY 288 acatactgctgtcttcacacaggaataatggtgctcttagactcttttggttcacaga 347
 Db 126299 ACATACCTGCTACTCTTCACTCAGGCGACTACTGGTGTCTCTCCGACTTATTGAAATCATGGA 126240
 QY 348 atactcag---gatgactcttagtataaggtcattgctctttttaaaacaagtgatgagg 404
 Db 126239 ATATTCAGACAATGATCTAAAGTGTAAAGTATCATCATTTTAAACAGGGTGATGAGG 126180
 QY 405 gctctcctcatctgcacccccctgctcctctgagtggtgtcccaggccat---catcagcccca 461
 Db 126179 GAGTCTCCATCTGCACACTTGGCTCTCTAAGCGTGTCTCAGGCGCATCACCATCAGCCCCA 126120

QY 462 gcattctctcttgccaaagctcaaacatctcttctgcaagtcacatctttaggattcttcc 521
 Db 126119 GCACCTCTCTTCTGGAAGTTTAAACATATTCTGCAAAATCACACCTTAGCTTCATCC 126060
 QY 522 ttttctcatgggtctcctcaacatgttctcattggtg-aaatctctgctgtacactgcgacta 580
 Db 126059 TCTTCTCATGGTCTCTCAACATGTTTCTAATAACCTTTTCTGCTTTTCTATTGCTGCTA 126000
 QY 581 cccccagt-gaaacggggcagctctcttctgtctacacagcactgtctcttttgcaca 639
 Db 125999 CCGCAATAGATTGGGGCCAGTCTTTTGTGTCACCTGAGCACTGTATGTTTGGCCA 125940
 QY 640 tgagctacacccacagagactgtttttcacactaatgactttgaggatgtcaccttta 699
 Db 125939 TGAGCTACACCCACAGAGGCTGTTTTCATCTAATAATGTTTTCAGGGATGTCATCTTCA 125880
 QY 700 taggtttcaggtctcctcctcaagagctacatggtgattattttatcacagacaataagagg 759
 Db 125879 TAGGACTCATGGTCTCTCAAGTGGCTA-TGGGTGATATTATTTGTACAAACA-AAGAGA 125822
 QY 760 ctatctcagtgcccttcacacagccagcctgtccccc--gagctccaccagtgaaaagagcc 817
 Db 125821 CTGTCCACGACCTACATAGCACCGCGTCTCCCAAGAGCTCTCCCGAGCAAAAGGGCC 125762
 QY 818 tccaggtctatcttactcgtcgtgga---ggttctctccacatactggtggagactttacg 874
 Db 125761 TCCAGACCATCTCTATTGTTGGTGAGCTGTTTGTCTTCATATATATTGGTGGAGCTTCATA 125702
 QY 875 ttctcatttcagaggtgtgacatgataaatgattctcgtctgtagtggctccaggtt 934
 Db 125701 TTCTCATTTTTCAGTGGTGTGACATGGATAACGACCTCTACGCGTATGATTCCATATG 125642
 QY 935 attgtggcaatagctatgcgcgaattagtccttcttgatcctaatttctgctgataaccaa 994
 Db 125641 CCTGTGCTCAATAACTATGCCATAATTAGTCCTTAGCTTACCTATGCTGATAA-CAA 125583
 QY 995 atattcaagactctgcaaatgttatgg 1021
 Db 125582 ATATTCAAGATTGTTTCAGACTCTTTGG 125556

RESULT 5

AC104335 217087 bp DNA linear HTG 08-DEC-2001
 LOCUS Homo sapiens chromosome 1 clone RP11-433K2, WORKING DRAFT SEQUENCE,
 DEFINITION 6 unordered pieces.
 AC104335 AL359674
 AC104335.1 GI:17426299
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 217087)
 AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 217087)
 AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Submitted (08-DEC-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 COMMENT On Dec 8, 2001 this sequence version replaced gi:9801069.
 ----- Genome Center
 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: <http://www.genome.washington.edu>
 Contact: uwgchgs@u.washington.edu
 Drafting Center: SC
 ----- Project Information
 Center project name: chr-1
 Center clone name: RP11-433K2 (sc0158)
 ----- Summary Statistics

Sequencing vector: plasmid; 50% of reads
Chemistry: Dye-terminator ET; 64% of reads
Chemistry: Dye-terminator Big Dye; 36% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 213380 bases at least Q40
Consensus quality: 215084 bases at least Q30
Consensus quality: 216045 bases at least Q20
Insert size: 256066; 17.2% error; agarose-fp
Insert size: 216587; sum-of-contigs
Quality coverage: 6.3x in Q20 bases; agarose-fp
Quality coverage: 7.5x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 4074: contig of 4074 bp in length
* 4075 4174: gap of unknown length
* 4175 9708: contig of 5534 bp in length
* 9709 9808: gap of unknown length
* 9809 13642: contig of 3834 bp in length
* 13643 13742: gap of unknown length
* 13743 52846: contig of 39104 bp in length
* 52847 52946: gap of unknown length
* 52947 140430: contig of 87484 bp in length
* 140431 140531: gap of unknown length
* 140531 217087: contig of 76557 bp in length.

FEATURES

source

1. 217087
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-433k2"

/library="RP11-433k2"
/library="RP11-433k2"

misc_feature

1. 4074

misc_feature

4175..9708

misc_feature

9809..13642

misc_feature

13743..52846

misc_feature

52947..140430

misc_feature

140531..217087

BASE COUNT

60435 a 48981 c 48285 g 58724 t 662 others

ORIGIN

Query Match

Best Local Similarity 33.78; Score 375.6; DB 2; Length 217087;

Matches 547; Conservative 0; Mismatches 134; Indels 21; Gaps 7;

Y 1 attccagatcatagaga-----tggtgaattggtttattgagacatggcagaaa--- 53

Db 139231 ATCCAGATCAGACGATTTATGAACATAATGATACATAATTAACACATGCGCAGAAATGT 139290

Y 54 ---ttatgctattctattagatctctgtctttctccacagatctcttgcttaatt 110

Db 139291 TCCTAATGCAATATATCTTATGCTCTTATTTCTCTACAGATATCTTTATTTGTTT 139350

Y 111 ttccttcttaag---atgacacaaactctggtttttattacacataaattctttatc 167

Db 139351 AATTTTCATGGAATGACCAAGCTTTTCAGCTATATGTCATTTAAAGATGTTATTACC 139410

Y 168 cacaagcagctttggaatttcagacaacacacatctcttctttccacatcttcacct 227

Db 139411 CTCAGTCAGCTTTGGAAATCTCAGCAAAACACCTTCCTCTCTCTTTTCCACATCTTCACCT 139470
Y 228 ttgttttcagtcacaggtcttaagtcctattgacatgataaattagtcacactgtctctcatcc 287
Db 139471 TTGCTTATACTACACAGGCTTAAGCCCATTTGACATGACCATAGTACCTAGTGCCTTAATCC 139530
Y 288 acatactgctgctctcttcactcaggcaaatattggtgctcttagaacttttttggttcaaga 347
Db 139531 ACATACCTGCTACTCTTCTCAGTCAGGCGATACCTGGTCTCTCCGACTTATTTGAATCATGGA 139590
Y 348 atactcag---gatgacttaggtataaggtcattgtctcttttaaaagaaggtgatgagg 404
Db 139591 ATATTCAGAACATGATCTAAGGTGTAAGATCATCACATTTTAAACAGGGGTGATGAGG 139650
Y 405 gctctccatctgcacccctctgctctgagtgctgctcagggccat---catcagcccca 461
Db 139651 GAGTCTCCATCTGCACCACTTGCCTTAAAGCGTCTCCAGGCCATACCATCAGCCCA 139710
Y 462 gcatctctcttgagcaagctcaaacatctcttgcaagtcacatcttagagattcttcc 521
Db 139711 GCACCTCTCTTCTGGAAGGTTTAAACATATTTCTGCAAAATCACACCTTAGGCTTCATCC 139770
Y 522 ttctctcatgggtcctcaacatgttcattggt-gtaactctctgctgtacactcgggcta 580
Db 139771 TCTTCTCATGGGCTCTCAACATGTTCTACTAATAACCTTTTGTCTTTTCATTTGTCCTA 139830
Y 581 ccccccagtg-gaaacggggcagctctctgtttgtctacacagcactgttctcttttgccca 639
Db 139831 CCCGGAATAGGATTTGGGGCCAGTCTTTTGTGTCACCTGAGCAGCTGTTATGTTTGGCCA 139890
Y 640 tgagctacacacacagagagactgtttttcacactaaagactt 681
Db 139891 TGAGCTACACCCAGGAGGCTGTTTTCATACCTAAAGCCCTT 139932

RESULT 6

AC074050/c

LOCUS

AC074050

DEFINITION

AC074050

AC074050.3

GI:18057079

HTG.

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 73661)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished

2 (bases 1 to 73661)

DOE Joint Genome Institute.

Direct Submission

Submitted (13-JUL-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 73661)

DOE Joint Genome Institute.

Direct Submission

Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Jan 4, 2002 this sequence version replaced gi:15963644.

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

genbank flat file format but are available as part

of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality

Oy	784	agcctgtcccc--gagtctcaccaagtgaagaagagcctccaggctatcttacttgttggtg	841
D6	72858	CACCTCTCCCAAGACCCTCCCAGAGAAAAAGGGCCACCCAGACCATCCTGCTGTGTGTG	72799
Oy	842	agt----ttgtcttcacatactggtagctttaagcttctcatcttcaggagagtgtagaca	898

Qy	899	tggaataatgattctc-----tgcattgtggctccagggtattgttgccaatagctatgcc	955
Db	72738		
Qy	956	gcaattagctcttgatgctctaattattactgataaccacaataattcaagaactctgcaaa	1013
Db	72678		
RESULT	7		
AC106730		152748 bp	DNA
LOCUS			linear
DEFINITION		152748 bp	PRI 03-FEB-2014E2, complete sequence.
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			

[illegible]

8	RESULT	AC073188/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	REFERENCE								
	AUTHORS								
	TITLE								
	JOURNAL								
	MEDLINE								
	REFERENCE								
	AUTHORS								
	TITLE								
	JOURNAL								
	REFERENCE								
	AUTHORS								
	TITLE								
	JOURNAL								
	REFERENCE								
	AUTHORS								
	TITLE								
	JOURNAL								
	REFERENCE								
	AUTHORS								
	TITLE								
	JOURNAL								
	COMMENT								

AC073188 55128 bp DNA linear PRI 09-JAN-2002
Homo sapiens BAC clone RP11-73B2 from 7, complete sequence.
AC073188
AC073188.10 GI:15920117
HTG.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 65128)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
2 (bases 1 to 65128)
Cordum, H. and Abbott, A.
The sequence of Homo sapiens BAC clone RP11-73B2
Unpublished (2001)
3 (bases 1 to 65128)
Waterston, R.H.
Direct Submission
Submitted (10-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 65128)
Waterston, R.H.
Direct Submission
Submitted (04-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 65128)
Waterston, R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 4, 2001 this sequence version replaced gi:14647308.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: saplens@waton.wustl.edu
----- Summary Statistics

Center project name: H_NH0073B02

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7/>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/qsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male

donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frenken, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-905H7, 2000 bp overlap; the clone sequenced to the right is RP5-1193P9, 2000 bp overlap. Actual start of this clone is at base position 36351 of RP11-73B2; actual end is at base position 72100 of RP5-1193P9.

Polymorphisms exist between AC073188 and AC006457. A single plasmid region exists between 68464-68496 and 68531-68559. A single stranded/chemistry region exists at base 68546. This clone contains an extra sequence fragment not represented in the finished neighbor, AC006457. This is submitted as H_NH0073B02_F1.

FEATURES

```

source
1..65128
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7"
/clone="RP11-73B2"
406..1310
/rpt_family="L1"
1317..1547
/rpt_family="L1"
3257..3634
/rpt_family="L1"
3635..3925
/rpt_family="MaLR"
3930..4120
/rpt_family="MaLR"
4136..4334
/rpt_family="Alu"
4341..4488
/rpt_family="Alu"
4555..4753
/rpt_family="L1"
5597..5693
/rpt_family="MER1_type"
5802..6205
/rpt_family="L1"
9188..9281
/rpt_family="MER53"
10528..10758
/notes="similar to EST AI028332 (NID:g3245641) ov90g03.x1"
10576..10799
/rpt_family="MaLR"
10813..11168
/rpt_family="MaLR"
11452..11647
/rpt_family="MER2_type"
11646..11696
/rpt_family="(TATG)n"
11701..11833
/rpt_family="Alu"
12521..13095
/notes="similar to EST AI806860 (NID:g59393426) wf36b07.x1"
13058..13087
/rpt_family="AT_rich"
13146..13198
/rpt_family="AT_rich"
13365..13671
/rpt_family="Alu"
14087..14115
/rpt_family="(TTTTTA)n"
14129..14883
/rpt_family="L1"

```

```

repeat_region 15223..15246
/rpt_family="AT_rich"
misc_feature 15290..15571
/notes="similar to EST AW445087 (NID:g6986849)"
misc_feature 15290..15571
/notes="similar to EST AW627961 (NID:g7374751) hi36dl2.x1"
repeat_region 16305..16574
/rpt_family="Alu"
16753..17021
/rpt_family="Alu"
17038..17286
/rpt_family="Alu"
17700..18467
/rpt_family="L1"
17853..17872
/rpt_family="(CAAAA)n"
18468..18778
/rpt_family="Alu"
18752..18796
/rpt_family="(A)n"
18779..18938
/rpt_family="L1"
18838..18870
/rpt_family="AT_rich"
18944..19246
/rpt_family="Alu"
19220..19265
/rpt_family="(GAAA)n"
19256..19691
/rpt_family="ERV1"
19748..19790
/rpt_family="(CAAAA)n"
20557..20613
/notes="similar to EST AW445087 (NID:g6986849)"
20557..20666
/notes="similar to EST AI028332 (NID:g3245641) ov90g03.x1"
20594..20828
/notes="similar to EST AA292895 (NID:g1941876) zt66c08.r1"
21125..21428
/rpt_family="Alu"
21664..21691
/rpt_family="AT_rich"
22442..22750
/notes="similar to EST BF899282 (NID:g12290741)"
22640..23050
/notes="similar to EST AI279713 (NID:g3917947) ql53f03.x1"
22855..23086
/notes="similar to EST AA292895 (NID:g1941876) zt66c08.r1"
23402..23451
/rpt_family="T-rich"
23430..23733
/rpt_family="Alu"
23850..24148
/rpt_family="MER2_type"
24566..25150
/rpt_family="L1"
25151..25275

```

Query Match 27.3%; Score 304.6; DB 9; Length 65128;

Best Local Similarity 62.1%; Pred. No. 5.4e-68;

Matches 639; Conservative 0; Mismatches 364; Indels 26; Gaps 9;

QY 11 atagagatgttgaaatgggtattattgagaaacatggcagaaattatgcattctcatta 70

Db 37223 ATTAATAATGATAATATGGTAATAATTTTGACCGTGAATAATAATACTACTGATAGGA 37164

QY 71 gatctgtgttttctccacagatatcctttgttttaattttcttcttaagatcaaa 130

Db 37163 TATCTATTGCTTTTTCAGATATATAAATATGATCCCTCTTACCATTAAAGACCATTA 37104

QY 131 ctctctgtgtttattaccacaaatcttctttttccacaaagccagcttttgggaattca 190

Db 37103 G-TCCACTTATGACATATTAAAAATTTGCCCTTTTTCCTCCCAAGCTGGTATTGGACTACA 37045

QY

371

aagggtcaattgtctctttttaaacaaagtgtagagggccctctctccatctgcacccccctgcttc

430

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

QY 714 ctctcaagaggtactaggtgattattttatcacagacaataaagaggctatctcagtgct 773
Db 601 CTCTCAAGTGGATACATGGTGACTCTCTGTACAGGC-ATAAGAGGCAGACCCAGCATCT 659
QY 774 tcacacagcagcctgtccccc--gagctcacacagtgaaagagcctccacaggctatctt 831
Db 660 TCAGAGCACCAGCGCTTCTCCAAAAGCATCCCAAGAAAGAGGGCCACCCAGACCATCT 719
QY 832 actctggtgagt--tttgtcttcacatactgggtgagctttacagttctcattttcagg 888
Db 720 ATTCTCTCATGAGTCTCTTTGGTGCATGTACTTTTGGACATAACTGTCTCTGGTCTC 779
QY 889 aggtgtgacatggataaattgattctctgctagtgctgctccaggttattgtggccaatag 948
Db 780 AAGATGTGTGGGATATCGACTCTGTTCGTATGTGTGTCAGATGCTTGTGGGAATGG 839
QY 949 ctatgcgcgaatagctctttgctgacttaattatgctgataacacaataattcagaactct 1008
Db 840 CTATGCCAGCATCAGCCCTTTGGTGTCTAATCAGTACTGAAAAGAGAAATAATCAAGGTCTT 899
QY 1009 gcaaatgttatggtttaat 1028
Db 900 AAAATACATATGGGTGAGAT 919

RESULT 11
AC108131 151941 bp DNA linear PRI 06-FEB-2002
LOCUS AC108131 Homo sapiens chromosome 16 clone RP11-337N9, complete sequence.
DEFINITION AC108131 AC022860
ACCESSION AC108131.2 GI:18542984
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 6, 2002 this sequence version replaced gi:18369977.
Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES
source Location/Qualifiers
1. .151941

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-337N9"
BASE COUNT 42554 a 32907 c 31887 g 44593 t
ORIGIN
Query Match 25.2%; Score 281.2; DB 9; Length 151941;
Best Local Similarity 60.4%; Pred. No. 6.6e-62;
Matches 568; Conservative 0; Mismatches 363; Indels 9; Gaps 6;
QY 11 atagagagtggtgaattggttattattgagaacatgagcagaataattgctattctcatta 70
Db 70064 ATTAAATGATAAGGTTGGTAATAATTTTGACTATACAATAACATAATTCATGATGA 70123
QY 71 gatctctgtctttctccacagatat--ccttggcttttaattttctctcttaagatgataa 129
Db 70124 GGTCTATTGCTTTTTCAGGTATATAAAATATAATTCCTCTTCATCATTTAAGACGACCA 70183
QY 130 actctcctgggttttattaccatacaaaatctcttttaccacaaagccagctttgaaattc 189
Db 70184 TGTTCAGTTAAGGTACTACTAACAAATGGCCCTCTTTTTCAGCTGTGTAFTGACCTAG 70243
QY 190 agcaaacaccatcctctcttctttccacatcttcacacctttgttttcagtcacaggtctaa 249
Db 70244 AGCCAACAACCTCCTCTATTTTCCAAATCTTCTCACTCCTTCAGGATCACAGGCCTAA 70303
QY 250 gtccattgacatgataattagtcacctgtctctcattccacatactgtctgtctctcaactca 309
Db 70304 GCCCACTGACCCCATCACCTGTCACTGGCCCTTGTCCACTTAGGGATGTCTCCTCATTTG 70363
QY 310 ggcaaatattggtgccttagactcttgggttcacagaataactcagatgactcttagta 369
Db 70364 GGTCTCTTGGCATCTCCAGACCTGTTTGAGTCACTGTATTTTCAGAAATGACTTCAAGTG 70423
QY 370 taaggctcattgtctttttaacaagggtgtagggggcctctccatctggttccctccctccct 429
Db 70424 TAAGGCATTCTCTGTCATGCACAGGGTGATGAGGAGCTCTCCATCTGCACACCATGCT 70483
QY 430 cctgagtgctccagc---catcatcagccacagactctctccttgcacaaagctcaa 486
Db 70484 CCTGAGCATGCTCCAGGCTGTGCGCATCAGCCCTGGCCACTCTCTGGTCTGCAAGGATTA 70543
QY 487 acatcctcttgcaagtcacatctttagattctctctttctctcattgggttccctccacagt- 545
Db 70544 ACAGAAATTCAAAGGTTACATCTTTCACTCCTCTCTTCTCTATGGGTCTCTCAGTTTGT 70603
QY 546 tcattggtgtaattctctgtgtacactggggtacacccacagtgaaacggggccagtcctt 605
Db 70604 TCTCAGTAGTAACCTGCTCTCCTCCACTGTGGCTTCTTAATGGGACCAAGAGCGTTGT 70663
QY 606 -ctgtttgtcatcacagcactgttcccttttggccatgagctacacccacaggaagactgtt 664
Db 70664 GCTAAGTATCAGTAAATACTGATCACTTTCTTCATTAAGCTATCATTAAGGAGCCCTGTC 70723
QY 665 ttccacactaatgactttgagggtgctcacctttatagggttccatggttccctcctcaagagg 724
Db 70724 TTTTCATGCTGCCACTGCTTACGAATGTCTCTTTGTAGCAATCATGCAGGCCCTCAAGTGC 70783
QY 725 ctacatggtgattattttatcacagacaataagaggtctatctcagtgcttccacacagcca 784
Db 70784 ATACATGGTAATTTCTCTTTGTTCAGGC-ATCAGAGACAATCCCAATACCTTTTCATAGAACCA 70842
QY 785 gcctgtcccgagctcaccagtgaaagagcctcccgagctatcttactgctggtgagt 844
Db 70843 ACCTCTCTCAAGAGCCTCCCCAGAGAAAGGGGACCA--TCCTGTTGCTGGTGAGTTGC 70900
QY 845 ttgtcttcacatactgggtggactttacgttctctcatttccaggaggtgtgacatggata 904
Db 70901 TTTCTGTCATGTACTGGGGGACCTTATTATCTCATCTCCTCCTCAACTCTATTATTGGCT 70960
QY 905 aatgattctctgctagtgctggctccaggttattgtggcca 944

Db 148141 GGCTTCTTGGCATCTCCAGACCTGTTTGGTCACTGATTTTCAGAAATGACTTCAAGTG 148082
QY 370 taaggtcatgtctcttttaacaaggtgatgaggccctccatctgcacccctgcct 429
Db 148081 TAAGGCATTTCTTGCATGACAGAGGTGATGAGAGCTCTCCATCTGCACCCCTGCCT 148022
QY 430 cctgagtgctccaggc---catcatcagccccagcatctctccttggcacaagctcaa 486
Db 148021 CCTGAGCATGCTCCAGGCTGTCGCCATCAGCCCTGGCACCTCCTGGTCTCGAAGGATTAA 147962
QY 487 acatcctctgcgaagtcaacatcttaggattctctctcttctctcttggtctcacaatgt- 545
Db 147961 ACAGAAATTCAAAAGTTTACATCTTTCACCTCTTCTTCTCTCTATGGTTCAGTTTGTG 147902
QY 546 tcattggtgtaactctctgctgtacactgcgctacccccagtgaaacgggcccagctctt 605
Db 147901 TCTCAGTAGTAACCTGCTCTCCACCTGTGGCTTCTTCTTAATGGGACCAAGAGCGTGT 147842
QY 606 -ctgtttgtcatcagcactgttctcttcttggccatgagctacacacacagggactgtt 664
Db 147841 GCTAAGTATCAGTAATACTGATGATCACTTCTTCCATAAGCTACATTTATCAGGAGCTGTC 147782
QY 665 ttccacaactaagacttgaggatgtcacactttatagggttcatggtctctctcaagagg 724
Db 147781 TTTCACTGCTGCCACTGCTTACGAATGCTTCTTGTGTAGCAATCATGCGCCCTCAAGTGC 147722
QY 725 ctacatggtgattattttatcacagacaataagaggtctatctcagtgctcttcacacagcca 784
Db 147721 ATACATGTAATCTCTTGTTCAGGC-ATCAGAGACAATCCCACTACCTTCATAGAACCA 147663
QY 785 gctgtccccgagctctcacagtgaaagagctcccaagctctcttactgctggtgagt 844
Db 147662 ACTCTCTCCAGAGGCTCCCCAGAGAAAAGGGGACCA--TCTTGTGTGCTGGTCAAGTGC 147605
QY 845 ttgtcttcacactggtggtgactttacgttctctcatttcagagaggtgtgacatggata 904
Db 147604 TTCTGCTCATGTACTGCGGGGACCTTATTCATCTCCTCCCACTCTATTATTGGCT 147545
QY 905 aatgattctgctagtggtggtccaggttattgtggcca 944
Db 147544 TATGACTCAGTCATGTTGTGAGCATGCGAGGCTTCTGGCCA 147505

RESULT 13
AC022202/c
LOCUS
DEFINITION Homo sapiens BAC clone RP11-321E8 from 7, complete sequence.
ACCESSION AC022202
VERSION AC022202.12 GI:13992747
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 95097)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 95097)
AUTHORS Meyer,R., Abbott,A. and Hawkins,M.
TITLE The sequence of Homo sapiens BAC clone RP11-321E8
JOURNAL Unpublished
AUTHORS Waterston,R.H.
REFERENCE 3 (bases 1 to 95097)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 95097)
AUTHORS Waterston,R.H.
TITLE Direct Submission

Submitted (08-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 95097)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 8, 2001 this sequence version replaced gi:13431144.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0321E08

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 17 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR17, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-468B6, 200 bp overlap; the
clone sequenced to the right is RP11-561N12. Actual start of this
clone is at base position 61428 of RP11-468B6; actual end is at
base position 95097 of RP11-321E8.

FEATURES

source
1. .95097
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="7"
/clone="RP11-321E8"
/clone_lib="RPCI-11"
repeat_region
273. .564
/rpt_family="Alu"
repeat_region
1280. .1352
/rpt_family="L2"
repeat_region
1371. .1404
/rpt_family="MER1_type"
repeat_region
1570. .1808
/rpt_family="L2"
repeat_region
1861. .2172


```

Db 35980 GAATGTGAGTAAATACCTGCTCACTTTCCCAATGAATGTCAACATCAGAAAGGCTGTTGT 35921
Qy 668 cacactaatgacttgaggatgcacacttataggttcattggtctctcaagagccta 727
Db 35920 TACTCTGCTGTTATCCAGAGATGCTTCTTCCCTAGGAAATCATGCCGCTCTCAAGTGCCTA 35861
Qy 728 catggtattattttacacagacaataagaggtctatctcagtcgcttcacacagccagcc 787
Db 35860 CATGGTATCTCTGTGCCAGGC-ATCAGAGGCGCTCCAGCAGCCTTCACAGCACTAGCT 35802
Qy 788 tgt--ccccagtgctaccagtgtaaaagagcctccagggctattctactctggtgagt- 844
Db 35801 TTTTATTAAAGAACCTCCCCAGAGAAAGAGGCGCCACCAAGACCATCTTGCTGCTGGTAGTT 35742
Qy 845 -tttgcctcacatactggtgagactttacgtcttcatttccagggaggtgacatgga 902
Db 35741 TCCTTGGTGTATGACTCATTTGACCTTAATGTTCTCATCTCCACAAATGTTGTATGGG 35682
Qy 903 taaatgattctgctgtdgtggtcctcaggttattgtggccaatagctatgcgcgaatta 962
Db 35681 TATTACGCCCTGTCATCTACAGTGTCCACAAAGTTTATGGTCAATGCCCTATGCCACTGTCA 35622
Qy 963 gtcccttgatgctaattatgctgataaccaaatattcaagagactctgcaaatgtt 1017
Db 35621 GTCCATGGTGTCTAATCAGATCTGATAAAAGAAATCATCACTATTCTGCCAAAGGT 35567

RESULT 14
AC073188
LOCUS AC073188 65128 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-73B2 from 7, complete sequence.
ACCESSION AC073188
VERSION AC073188.10 GI:15920117
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 65128)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 65128)
AUTHORS Cordum,H. and Abbott,A.
TITLE The sequence of Homo sapiens BAC clone RP11-73B2
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 65128)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 65128)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 65128)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 4, 2001 this sequence version replaced gi:14647308.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0073B02

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mail to: sgreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-905H7, 2000 bp overlap; the clone sequenced to the right is RP5-1193P9, 2000 bp overlap. Actual start of this clone is at base position 36351 of RP11-73B2; actual end is at base position 72100 of RP5-1193P9.

Polymorphisms exist between AC073188 and AC006457. A single plasmid region exists between 68464-68496 and 68531-68559. A single stranded/chemistry region exists at base 68546. This clone contains an extra sequence fragment not represented in the finished neighbor, AC006457. This is submitted as H_NH0073B02_F1.

FEATURES

FEATURES	Location/Qualifiers	source
	1..65128	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="7"	
	/map="7"	
	/clone="RP11-73B2"	
	/clone_lib="RPCI-11"	
repeat_region	406..1310	
	/rpt_family="L1"	
repeat_region	1317..1547	
	/rpt_family="L1"	
repeat_region	3257..3634	
	/rpt_family="L1"	
repeat_region	3635..3925	
	/rpt_family="MaLR"	
repeat_region	3930..4120	
	/rpt_family="MaLR"	
repeat_region	4136..4334	
	/rpt_family="Alu"	
repeat_region	4341..4488	
	/rpt_family="Alu"	
repeat_region	4555..4753	
	/rpt_family="L1"	
repeat_region	5597..5693	
	/rpt_family="MERL_type"	
repeat_region	5802..6205	


```
repeat_region /rpt_family="L1"
9188..9281
/rpt_family="MER53"
10528..10758
/note="similar to EST AI028332 (NID:g3245641) ov90g03.x1"
repeat_region /rpt_family="MaLR"
10576..10799
/rpt_family="MaLR"
10813..11168
/rpt_family="MaLR"
11452..11647
/rpt_family="MER2_type"
11646..11696
/rpt_family="(TATG)n"
11701..11833
/rpt_family="Alu"
12521..13095
/note="similar to EST AI806860 (NID:g5393426) wf36b07.x1"
repeat_region /rpt_family="AT_rich"
13058..13087
/rpt_family="AT_rich"
13146..13198
/rpt_family="AT_rich"
13365..13671
/rpt_family="Alu"
14087..14115
/rpt_family="(TTTTTA)n"
14129..14883
/rpt_family="L1"
15223..15246
/rpt_family="AT_rich"
15290..15571
/note="similar to EST AW445087 (NID:g6986849)"
15290..15571
/note="similar to EST AW627961 (NID:g7374751) hi36d12.x1"
16305..16574
/rpt_family="Alu"
16753..17021
/rpt_family="Alu"
17038..17286
/rpt_family="Alu"
17700..18467
/rpt_family="L1"
17853..17872
/rpt_family="(CAAAA)n"
18468..18778
/rpt_family="Alu"
18752..18796
/rpt_family="(A)n"
18779..18938
/rpt_family="L1"
18838..18870
/rpt_family="AT_rich"
18944..19246
/rpt_family="Alu"
19220..19265
/rpt_family="(GAAA)n"
19256..19691
/rpt_family="ERV_L"
19748..19790
/rpt_family="(CAAAA)n"
20557..20613
/note="similar to EST AW445087 (NID:g6986849)"
20557..20666
/note="similar to EST AI028332 (NID:g3245641) ov90g03.x1"
20594..20828
/note="similar to EST AA292895 (NID:g1941876) zt66c08.r1"
21125..21428
/rpt_family="Alu"
21664..21691
/rpt_family="AT_rich"
22442..22750
/note="similar to EST BF899282 (NID:g12290741)"
22640..23050
/note="similar to EST AI279713 (NID:g3917947) ql53f03.x1"
```

```
misc_feature 22855..23086
/note="similar to EST AA292895 (NID:g1941876) zt66c08.r1"
repeat_region /rpt_family="T-rich"
23402..23451
repeat_region /rpt_family="Alu"
23430..23733
23850..24148
/rpt_family="MER2_type"
24566..25150
/rpt_family="L1"
25151..25275

Query Match 25.1%; Score 279.6; DB 9; Length 65128;
Best Local Similarity 62.6%; Pred. No. 1.6e-61;
Matches 553; Conservative 0; Mismatches 319; Indels 12; Gaps 7;

Qy 140 ttattaccatacaaaattctttttatccacaagccagcttggtaatttcagcaaacacc 199
Db 12154 TGTATGTTATTTAAAAGCTGCTTTTGTACAGTTGCTGTGGAATCTCAGTCAATACT 12213
Qy 200 atccttctttttccacatcttccaccttggtttttcagtcacaggtcttaagtcattgac 259
Db 12214 TTTCTGTCTCTTTCAGCATCTTCACACTCTCTCTGTATACAGACTAAACCCACACTGAC 12273
Qy 260 atgataatagtcacctgtctctcatccacatactgtctcttcacactcagcgcaatttg 319
Db 12274 TTGATCATCTGTCCTTGGCCCTCGTCCACATAGTGAACCTTCTTCATTTGCATGTTATTG 12333
Qy 320 gtgtccttagactcttcttggtttcacagaatactcagatgatcttaggtataaagtcatt 379
Db 12334 CAGTCTCCAAATCTGTTTGCATCATTTGAA-CTTCAGAATGACTTCAAGTGTAAAGCACTT 12392
Qy 380 gtctttttaacaagggtgatgaggggcctctccatctcgcacccctgcctcctcctcagtg 439
Db 12393 TTCCACCTGAGCAGGGTGTGTGGGGCTCTCTATCTGCACACCTGCCCTCAGTGTG 12452
Qy 440 ctccagg---ccatcatcagcccgagcatcttctctctggcaagctcaaacatcctct 496
Db 12453 TTCAGGCCACCAACATCAGCCCCAGCACCTCTGATGGCAAGGTTTAAACATAAATCT 12512
Qy 497 gcaagtcacatcttaggattctctcttctctctgggttcctcctcaacatggttccatgtgt 555
Db 12513 TCAAAATTACATTTATCCATGTTTCTTCTTCTTATGTTGTTCTTCAACTTGTCTTTCAGTAGT 12572
Qy 556 aatcttctgtgtacactgagggtacccccagtg-aaacggggggcagctctctctgttctgc 614
Db 12573 AACAGGATCTTCTACACTGTAGTCTGCTTCCAGTGTGAGCCAGACAAATGTACTGAATATA 12632
Qy 615 atacgacactgtctcttcttggccatgagctacacccacagagagactgtttttcacacta 674
Db 12633 GGTAAAGTACTGCTCAGTGTCTTCTCCATGAGCTTCATCTCAGGGGCCACGTTTCTTATTCTG 12692
Qy 675 atgactttgagggtatgtcacctttatagggttccatgggttcctctcaagagggtacatggtg 734
Db 12693 ACAACACTTAGGGATGCTTTCTTGTAGGAATCTGCTGTTCTTAAATTGCATACATGTTG 12752
Qy 735 attattttatacagacaataaaggctatctcagtgcccttcacacagccagcctgtccccc 794
Db 12753 ACTCTCTATTTCAGGC-CTCAGAGGTGATCCACACCTTTCACAGCAATGTTCTCTTCCCC 12811
Qy 795 --gagctcacagtgaaagagctccaggctctctctcttactctgctggtgagttt--tgt 849
Db 12812 AGGAGTTCCTCCATGAGAAACGGGGCATCCAGACCATCTGCTGCTGGTGAAGTTTCTCTGT 12871
Qy 850 cttcacatactgggtggactttacgtttctctatcttttcagggaggtgtgacatggataaata 909
Db 12872 GGTCTCTGTTGGGTGGATTTTCATCATCTCTCTCTCACCANTGTTGTGGGCATATGG 12931
Qy 910 ttctctgtctag-tgtgggtccaggttatgttggtcccaatagctatgccgcaatagtccttt 969
Db 12932 CCCAATCAACCAGCAAGTTCAGAGTCTTGTAGTCAATGTCTATGCTGCGGTGATCAGTCTCT 12991
Qy 970 gatgctaatttatgctgataaccacaaatattcaagactctgcaaa 1013
```


Db 15886 ATTAAATGATAAGGTTGGTAATAATTTTGACTACACAAATACCTAACATTTGATGATAAGA 15945
Qy 71 gatctcttgcttttttccacagatat-cctttgtctttaatttctcttaagatgatcaa 129
Db 15946 GGTCTATTGCTTTTTCAGGTATATAAATAATTCCTCTTCATCATTAAGACGACCA 16005
Qy 130 acttctgggtttttattacacatacacaactctcttttatccacaagccagcttttgaatttc 189
Db 16006 TGTTCACAGTTAAGGTACTACTAACAAATTCGCCCTTCTTTTCAAGCTGGTATTGGACTCAG 16065
Qy 190 agcaaacaccatctcttcttttccacatcttccacctttgtttttcagtcacagagctcaa 249
Db 16066 AGCCAAACAATTCCTCCCTATTCTTCCAAATCTCTCACTCTCTCAGGATCACAGGCCTAA 16125
Qy 250 gtccattgacatgataattagtcacctgtctctctcatccacataactgtctcttctcaactca 309
Db 16126 GCCCAGTACGCCCATCACTCTGACCTGGCCCTTGTCCACTTAGGGATGCTCCTCATTTGT 16185
Qy 310 ggcaatatgtgtccttagactcttttgggttcacagaataactcaggatgattcttaggta 369
Db 16186 GGTCTTCTTGGCATTCTCAGACCTGTTTGTAGTCACTGTATTTTCAGAATGACTTCAAGTG 16245
Qy 370 taaggctcattgtctttttaacaagggtgatgggggctctccatctgcacccccctgcct 429
Db 16246 TAAGGCATTTCTTTCGATGACAGGGTGTGATGAGGAGCCTCTCCATCTGCACCACCTGCCT 16305
Qy 430 cctgagtgctccaggc---catcatcagccccagcatcttctctcttggcacaagctcaa 486
Db 16306 CCTGAGCATGCTCCAGGCTGTGGCATCAGCCCTGGCACCTCCTGGTCTGCAAGGATTAA 16365
Qy 487 acatccctctgcaagtcacatcttaggattctctcttctcttctcattgggtcctcaacatgt- 545
Db 16366 ACAGAAATTCAAAGGTTACATCTTTCACTCCTCTCTCTCTCTATGCGGTCTCAGTTTCTC 16425
Qy 546 tcattgggtgaatctctctgtctgtacactggcggtcaccgccagtgaaacggggccagctctt 605
Db 16426 TCTCAGTAGTAACCTGCTCTCTCCACTGTGCTCTTCTTAATGGGACCAAGACGGTTGT 16485
Qy 606 -ctgtttgctacacagactgttctctttttgcccattgagctacaccccacaggagactgtt 664
Db 16486 GCTAAGTATACAGTAAGTAATACATCACTTTCTTCCATAAGCTACATTTATCAGGAGCCTGTC 16545
Qy 665 ttccacacataatgactttgagggtgtgcacctttataggggttcattggtcctctcaagagg 724
Db 16546 TTTTCATGCTGCCACTGTTCGAATGTCTTCTTTGTAGCAATCATGACAGCCCTCAAGTGC 16605
Qy 725 ctacatgggtgattattttatcacacaataagaggctatctcagtgcccttcacacagcca 784
Db 16606 ATACATGGTAATTCTCTTTGTTTCAGGC-ATCAGAGACAAATCCCAATACCTTTCATAGACCA 16664
Qy 785 gcctgtccccagagtcacacagtgtaaaagagcctcccaggctatcttactctgtgtgagt 844
Db 16665 ACCTCTCTCCAAAGAGCCTCCCCAGAGAAAGGGGACCA--TCTTGTGCTGCTGAGTTGC 16722
Qy 845 ttgtcttcacatcaggtgaggttacgtcttcattcttccagaggtgtgacatgata 904
Db 16723 TTTCTGTCATGTACTGGGGGGAGCCTTATTAATCATCTCCTCTCAACTCTATTATTGGCT 16782
Qy 905 aatgattctctgtagtggtgctccaggattattgtggcca 944
Db 16783 TATGACTCAGTCATTGTGAGCATGCAGAGGCTTCTGGGCA 16822

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 01:51:05 ; Search time 107.09 Seconds
(without alignments)
2555.194 Million cell updates/sec

Title: US-09-728-309-1
Perfect score: 1114
Sequence: 1 attccagatcatagatgt.....gagaggttaatccattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62.4	5.6	7218	1	US-08-232-463-14
C 2	56.2	5.0	8353	1	Sequence 14, Appl
C 3	48.4	4.3	246240	2	Sequence 1, Appl
C 4	48.4	4.3	246240	2	Sequence 20, Appl
C 5	48.4	4.3	246240	2	Sequence 21, Appl
C 6	47.4	4.2	638	4	Sequence 22, Appl
C 7	44.4	4.0	497	4	Sequence 275, App
C 8	43.3	3.9	152331	3	Sequence 109, App
C 9	40.6	3.6	2270	2	Sequence 16, Appl
C 10	40.6	3.6	13674	2	Sequence 15, Appl
C 11	40.6	3.6	17410	1	Sequence 1, Appl
C 12	40.6	3.6	17410	1	Sequence 3, Appl
C 13	40.6	3.6	17410	1	Sequence 3, Appl
C 14	40.6	3.6	17410	2	Sequence 3, Appl
C 15	40.6	3.6	17410	2	Sequence 3, Appl
C 16	40.6	3.6	17415	3	Sequence 1, Appl
C 17	40.6	3.6	17415	5	Sequence 1, Appl
C 18	39.3	3.5	504	4	Sequence 120, App
C 19	36.2	3.2	80246	4	Sequence 4, Appl
C 20	36.2	3.2	80595	4	Sequence 3, Appl
C 21	35.4	3.2	289	4	Sequence 17, Appl
C 22	35.4	3.2	289	4	Sequence 17, Appl
C 23	35.2	3.2	2358	3	Sequence 1, Appl
C 24	35.2	3.2	2360	4	Sequence 10, Appl
C 25	35.2	3.2	40352	3	Sequence 15, Appl
C 26	34.2	3.1	4002	3	Sequence 9, Appl
C 27	33.6	3.0	437	4	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; CLONE: ptz9pt-F1s
; IMMEDIATE SOURCE:
US-08-232-463-14

Sequence 29, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 7, Appl
Sequence 6, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 6, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl


```
; NAME/KEY: misc.feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match 4.3%; Score 48.4; DB 2; Length 246240;
Best Local Similarity 83.3%; Pred. No. 0.0016;
Matches 55; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1011 aaatgttatggtttaaatatttctctctccaaagctcatgttgaaatttaacgccaat 1070
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 142977 AAGTGATATGTTGGATATTGTCCCTCTAAATCTCATGTTGAAATGTAATCTCCAAT 142918

QY 1071 gtggca 1076
||| |
DB 142917 GTTGGG 142912

RESULT 4
US-08-724-394A-21/c
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 01-OCT-1996
; APPLICATION NUMBER: US/08/724,394A
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-21

Query Match 4.3%; Score 48.4; DB 2; Length 246240;
Best Local Similarity 83.3%; Pred. No. 0.0016;
Matches 55; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1011 aaatgttatggtttaaatatttctctctccaaagctcatgttgaaatttaacgccaat 1070
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 142977 AAGTGATATGTTGGATATTGTCCCTCTAAATCTCATGTTGAAATGTAATCTCCAAT 142918

QY 1071 gtggca 1076
||| |
DB 142917 GTTGGG 142912

RESULT 5
US-08-724-394A-22/c
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 01-OCT-1996
; APPLICATION NUMBER: US/08/724,394A
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-22

Query Match 4.3%; Score 48.4; DB 2; Length 246240;
Best Local Similarity 83.3%; Pred. No. 0.0016;
Matches 55; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1011 aaatgttatggtttaaatatttctctctccaaagctcatgttgaaatttaacgccaat 1070
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 142977 AAGTGATATGTTGGATATTGTCCCTCTAAATCTCATGTTGAAATGTAATCTCCAAT 142918

QY 1071 gtggca 1076
||| |
DB 142917 GTTGGG 142912
```


Sequence 16, Application US/08852807
Patent No. 5861298
GENERAL INFORMATION:
APPLICANT: Debouck, Christine
APPLICANT: Drake, Fred
APPLICANT: Gowen, Maxine
APPLICANT: Rood, Julie
APPLICANT: Hastings, Gregg
APPLICANT: Adams, Mark
APPLICANT: Fraser, Claire
APPLICANT: Lee, No. 5861298man
APPLICANT: Kirkness, Ewen
APPLICANT: Blake, Judith
APPLICANT: Fitzgerald, Lisa
TITLE OF INVENTION: CATHEPSIN K GENE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,807
FILING DATE: 07-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,942
FILING DATE: 14-JUNE-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/020,273
FILING DATE: 17-JUNE-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/026,273
FILING DATE: 26-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50006-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2270 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-852-807-16

Query Match 3.6%; Score 40.6; DB 2; Length 2270;
Best Local Similarity 77.8%; Pred. No. 0.027;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1015 gttatggttttaaatattgtctctccaaagctcatgttgaatttaataatgcgaatgtgg 1074

Db 576 GATATGTTGGTGTGTGTGTCCTCCAGTCATCTGAAATGTGATCCCAATGTGG 517

Qy 1075 cag 1077

||

Db 516 GAG 514

RESULT 10
US-08-852-807-1/c
Sequence 1, Application US/08852807
Patent No. 5861298
GENERAL INFORMATION:
APPLICANT: Debouck, Christine
APPLICANT: Drake, Fred
APPLICANT: Gowen, Maxine
APPLICANT: Rood, Julie
APPLICANT: Hastings, Gregg
APPLICANT: Adams, Mark
APPLICANT: Fraser, Claire
APPLICANT: Lee, No. 5861298man
APPLICANT: Kirkness, Ewen
APPLICANT: Blake, Judith
APPLICANT: Fitzgerald, Lisa
TITLE OF INVENTION: CATHEPSIN K GENE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,807
FILING DATE: 07-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,942
FILING DATE: 14-JUNE-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/020,273
FILING DATE: 17-JUNE-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/026,273
FILING DATE: 26-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50006-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 13674 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-852-807-1

Query Match 3.6%; Score 40.6; DB 2; Length 13674;
Best Local Similarity 77.8%; Pred. No. 0.07;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1015 gttatggttttaaatattgtctctccaaagctcatgttgaatttaataatgcgaatgtgg 1074


```

;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
;
; US-07-841-646-3
;
; Query Match          3.6%; Score 40.6; DB 1; Length 17410;
; Best Local Similarity 77.8%; Pred. No. 0.079;
; Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
;
; Qy 1030 ttgtctctcctcaagctcatgttgaaatttaatcgcaatgtgagcagtaactaagaagtg 1089
;         ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8812 TTTGTCCCTACAAAACACTGTTGAAATTTAAATGTCAGTGTAACA/TATTGAGAGCTT 8871
;
; Qy 1090 atg 1092
;         |||
; Db 8872 atg 8874
;
; RESULT 12
; US-08-147-023-3
; Sequence 3, Application US/08147023
; Patent No. 5468845
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ERGIN
; APPLICANT: KUBERASMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,023
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 579,865
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; OTHER INFORMATION: THIS SEQUENCE."
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10696..10891
; OTHER INFORMATION: /label= EXON-2
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10960..10961
; OTHER INFORMATION: /label= GAP-2
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN
; OTHER INFORMATION: THIS SEQUENCE."
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11059..11211
; OTHER INFORMATION: /label= EXON-3
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11351..11352
; OTHER INFORMATION: /label= GAP-3
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
```

```

; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11420..11617
; OTHER INFORMATION: /label= EXON-4
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11721..11722
; OTHER INFORMATION: /label= GAP-4
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 13354..13436
; OTHER INFORMATION: /label= EXON-5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
; US-08-147-023-3
;
; Query Match 3.6%; Score 40.6; DB 1; Length 17410;
; Best Local Similarity 77.88; Pred. No. 0.079;
; Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
;
; Qy 1030 ttgtctctccaaagtcattgttgaaatttaacgcgaatgtgcagtgactaagaagtg 1089
;      ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8812 TTTGCCCTACAAAACACTGTTGAAATTAATTGCCAGTGTAACATTATTGAGAGGTT 8871
;
; Qy 1090 atg 1092
;      |||
; Db 8872 ATG 8874
;
; RESULT 13
; US-08-447-570-3
; Sequence 3, Application US/08447570
; Patent No. 5714589
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,570
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
;
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 579,865
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10696..10891
; OTHER INFORMATION: /label= EXON-2
; FEATURE:
```



```
;; NAME/KEY: exon
;; LOCATION: 3192..3730
;; OTHER INFORMATION: /label= EXON-1
;; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 10413..10414
;; OTHER INFORMATION: /label= GAP-1
;; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
;; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
;; OTHER INFORMATION: THIS SEQUENCE."
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: 10696..10891
;; OTHER INFORMATION: /label= EXON-2
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 10960..10961
;; OTHER INFORMATION: /label= GAP-2
;; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
;; OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN
;; OTHER INFORMATION: THIS SEQUENCE."
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: 11059..11211
;; OTHER INFORMATION: /label= EXON-3
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 11351..11352
;; OTHER INFORMATION: /label= GAP-3
;; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
;; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
;; OTHER INFORMATION: THIS SEQUENCE."
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: 11420..11617
;; OTHER INFORMATION: /label= EXON-4
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 11721..11722
;; OTHER INFORMATION: /label= GAP-4
;; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
;; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
;; OTHER INFORMATION: THIS SEQUENCE."
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: 13354..13436
;; OTHER INFORMATION: /label= EXON-5
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: 15044..15160
;; OTHER INFORMATION: /label= EXON-6
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: 17245..17410
;; OTHER INFORMATION: /label= EXON-7
;; US-08-449-700-3
```

```
Query Match 3.68; Score 40.6; DB 2; Length 17410;
Best Local Similarity 77.8; Pred. No. 0.079;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Caps 0;

Qy 1030 ttgtctctccaaagctcagtgtgaaatttaacgcgaagtgtggaagtactaagaagt 1089
      ||||| | | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8812 TTGTCCCTACAAAACATCATGTGAATTTAATTCGCACTGTAAACATTTATTGAGAGTT 8871

Qy 1090 atg 1092
      |||
Db - 8872 ATG 8874

RESULT 15
```

```
US-08-449-699A-3
; Sequence 3, Application.US/08449699A
; Patent No. 5958441
; GENERAL INFORMATION:
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: ANTIBODIES TO OSTEOGENIC PROTEINS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,699A
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/147,023
; FILING DATE: 01-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: STK-001CP6CN
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10696..10891
; OTHER INFORMATION: /label= EXON-2
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10960..10961
; OTHER INFORMATION: /label= GAP-2
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11059..11211
; OTHER INFORMATION: /label= EXON-3
```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11351..11352
OTHER INFORMATION: /label= GAP-3
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 11420..11617
OTHER INFORMATION: /label= EXON-4
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11721..11722
OTHER INFORMATION: /label= GAP-4
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 13354..13436
OTHER INFORMATION: /label= EXON-5
FEATURE:
NAME/KEY: exon
LOCATION: 15044..15160
OTHER INFORMATION: /label= EXON-6
FEATURE:
NAME/KEY: exon
LOCATION: 17245..17410
OTHER INFORMATION: /label= EXON-7
US-08-449-699A-3
```

```
Query Match 3.6%; Score 40.6; DB 2; Length 17410;
Best Local Similarity 77.8%; Pred. No. 0.079;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 1030 ttgtctctccaaagctcatgttgaaatttaacgcgaatggtgcagtaactaagaagtg 1089
||||| | | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8812 TTGTGCTCCCTACAAAACATCATGTTGAAATTTAATTGCCAGTGTACACATTATTGAGAGGTT 8871

Oy 1090 atg 1092
|||
Db 8872 ATG 8874
```

Search completed: July 21, 2002, 03:56:34
Job time: 7529 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 00:21:30 ; Search time 4107.22 Seconds
(without alignments)
5675.898 Million cell updates/sec

Title: US-09-728-309-1
Perfect score: 1114
Sequence: 1 attccagatcatagagatgt.....gagaggttaatccattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgO_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match Length	ID	Description
-----	-----	-----	-----	-----	-----

1	1096.4	98.4	206832	2	AL390728
c	1093.2	98.1	166798	30	AL390728
c	1093.2	98.1	217087	2	AC104335
c	541	48.6	206832	2	AL390728
c	375.6	33.7	217087	2	AC104335
c	311.6	28.0	73661	9	AC074050
c	311.6	28.0	152748	9	AC106730
c	303	27.2	65128	9	AC073188
c	303	27.2	139384	9	AC006455
c	302	27.1	989	4	AB064662
c	279.6	25.1	151941	9	AC108131
c	279.6	25.1	161609	9	AC011823
c	278.6	25.0	95097	9	AC022202
c	278	25.0	65128	9	AC073188
c	278	25.0	149496	30	AC022860
c	273.8	24.6	194142	9	AC092634
c	266.6	23.9	68299	9	AC091685
c	266.6	23.9	152964	2	AC011821
c	265.6	23.8	909	10	AY065469
c	264	23.7	909	10	AY065471
c	260.4	23.4	178105	9	AC073057
c	254	22.8	891	10	AY065468
c	250	22.4	903	10	AF291501
c	249.2	22.4	909	10	AY065466
c	248.2	22.3	900	10	AF291499
c	246.8	22.2	161571	9	AC069462
c	246.8	22.2	162339	2	AC024151
c	246.8	22.2	166991	9	AC090884
c	246	22.1	918	10	AY065467
c	245	22.0	1020	10	AY065473
c	243.2	21.8	188389	2	AC024199
c	243.2	21.8	195905	2	AC092447
c	242.8	21.8	894	10	AY065477
c	242.8	21.8	918	10	AY065470
c	242	21.7	202333	2	AC080144
c	241	21.6	957	10	AB062896
c	241	21.6	152964	2	AC011821
c	240.4	21.6	3403	10	MPHEREC2
c	239.6	21.5	912	10	AY065489
c	239	21.5	891	10	AY065474
c	238	21.4	1738	9	AF253312
c	236.4	21.2	897	10	AY065487
c	236	21.2	912	10	AY065488
c	235.8	21.2	900	10	AY065486
c	234.6	21.1	223275	2	AC098739

ALIGNMENTS

RESULT 1

AL390728
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-488L18, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.

ACCESSION AL390728

VERSION AL390728.22 GI:18614040

KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (sites)

Ashwell, R.

Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (05-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Feb 7, 2002 this sequence version replaced gi:18491329.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA488L18
 ----- Summary Statistics
 Sequencing program: XGAP4; version 4.5
 Chemistry: Dye-terminator; L08752; 100% of reads
 Dye-terminator: Big Dye; 90% of reads
 Consensus quality: 20499 bases at least Q40
 Consensus quality: 203211 bases at least Q30
 Consensus quality: 203559 bases at least Q20
 Insert size: 206732; sum-of-contigs
 Insert size: 205638; 4.3% error; agarose-fp
 Quality coverage: 9.91x in Q20 bases; sum-of-contigs Quality
 coverage: 10.88x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 204689: contig of 204689 bp in length
 * 204690 204789: gap of 100 bp
 * 204790 206832: contig of 2043 bp in length.
 Location/Qualifiers
 1. 206832
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-488L18"
 /clone_lib="RPC1-11.2"
 1. 204689
 /note="assembly fragment:03301"
 204790. 206832
 /note="assembly fragment:03801"
 BASE COUNT 58135 a 44577 c 44388 g 59631 t 101 others
 ORIGIN

Query Match 98.4%; Score 1096.4; DB 2; Length 206832;
 Best Local Similarity 99.6%; Pred. No. 1.8e-273;
 Matches 1110; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 1 attccagatcatagagatgttgaaattggtattattgagaacatggcagaaattatgct 60
 Db 144680 AFTCCAGATCATAGAGATGTTGAATTGGTTATTATTGAGAACATGGCAGAAATTATGCT 144739

QY 61 attcattagatctctgtgttttctccacagatatacttgccttaattttctctaa 120
 Db 144740 ATTCTCATTAGATCTCTGTGCTTTCTCCACAGATATCCTTTGCTTAATTTCTCTTCAA 144799

QY 121 gatgatcaacttctgtgtgttttattaccatacaaatctcttttattccacagcagctt 180
 Db 144800 GATGATCAAACTCTCTGTGTTTATTACCATCAAAATCTCTTTTATCCACAAAGCCAGCTT 144859

QY 181 tggatttcagcaaacaccatcttcttcttccacatcttccacttctgttttcagtc 240
 Db 144860 TGGAAATTCAGCAAAACACATCTCTCTTTCTCCACATCTTCACTTTGTTTCAGTCA 144919

QY 241 cagcttaagtccattgacatgataattagtcacctgtctctctccacatactgtgct 300
 Db 144920 CAGGCTAAAGTCCATTGACATGATAATTAGTCACTCTCTCTATCCACATCTCTGCT 144979

QY 301 ctccactcaggcaataattggtgctcttagactcttctgttccagataactcagatga 360
 Db 144980 CTTTCACATCAGCAATATGTTGCTCTTAGACTCTTTTGGTTTCAGAGAAATCTCAGGATGA 145039

QY 361 tcttaggataaaggtcattgtctttttaacaagggtgatggggcctctccatctcac 420
 Db 145040 TCTTAGGTATAAGGTCTATTGCTTTTAAACAAGGTGATGAGGGGCTCTCCATCTGCAC 145099

QY 421 cccctgctcctgagtggtgtgtccaggccatcatcagccccagcatcttctcttggcaaa 480
 Db 145100 CCCCTGCTCTCTGAGTGTGTCTCCAGGCGCATCATCAGCCCCAGCATCTCTCTTGGCAAA 145159

QY 481 gctcaaacatcctcttgcgaagtcacatcttagagattcttctcttctcagtggttctcaa 540
 Db 145160 GCTCAAAACATCCTTCTGCAAGTCACATCTTAGGATTCTTCTCTTTCTCATGGGTCTCAA 145219

QY 541 catgttctcattggtgtaattcttctgctgtacactgcgctacccccagtgaaacggggcca 600
 Db 145220 CATGTTTCATGGTGTAAATCTTCTGCTGTACACTGCGGCTACCCCGATGAAACGGGGCCA 145279

QY 601 gtcttctgtttgtacacacacactgttcttcttcttcttcttcttcttcttcttcttctt 660
 Db 145280 GTCTTCTGTTGTTCATACAGCACTGTTCTTTTGTGCCATGAGCTACACCCACAGGAGAC 145339

QY 661 tgttttccacactaatgactttgagggatgtcacttcttcttcttcttcttcttcttcttcttct 720
 Db 145340 TGTGTTTTCACACTAATGACTTTTCAGGGATGTCACCTTTATAGGGTTTCATGTTCTCTCAA 145399

QY 721 gaggtcacatgggtgattattttatcacagacaataaagggtctatctcagtccttccagca 780
 Db 145400 GAGGCTACATGGTGATTAATTTATACAGACAATAAGAGGCTATCTCAGTGCCTTCACACA 145459

QY 781 gccagcctgtccccgagctctcaccagtgaagagagctccccaggtcttcttcttcttctt 840
 Db 145460 GCCAGCCTGTCCCCGAGTCTCACCACTGAAAGAGGCTCTCCAGGCTATCTTACTGCTGT 145519

QY 841 gagtttcttctcactactggtggtgactttacgttctcttcttcttcttcttcttcttcttct 900
 Db 145520 GAGTTTGTCTTTCACATCTGCTGGTGGACTTTAGTTCCTATTTCAGGAGGTGTGACATG 145579

QY 901 gataaatgattctctgctagtggtgctccaggttatttggccaaatagctatgcgcgaat 960
 Db 145580 GATAAATGATTCTCTGCTAGTGTGCTCCAGGTATTGTGGCAATAGCTATGCCGCAAT 145639

QY 961 tagtcttctgctgaattatgctgataaccataattattcaagactctgcataatgttatg 1020
 Db 145640 TAGTCTTTGATGCTAATTTATGCTGTGAACCAATATTTCAAGAGCTCTGCAATGTATG 145699

QY 1021 gtttaaatatttctctctccaaagctcatgttgtaatttaataatcgaatgtggcaglac 1080
 Db 145700 GTTTAAATATTGTCTCTCCAAAGCTCATGTTGAAATTTAATCCCAATGTGGCAGTAC 145759

QY 1081 taagaagtgatgatgagagggttaataccatctcatg 1114
 Db 145760 TAAGAAG---TGATGAGAGGTTAATCCATTCATG 145790

RESULT 2

AL359674/c
 ID AL359674 standard; DNA; HTG; 166798 BP.
 XX AC AL359674;
 XX SV AL359674.4
 XX DT 20-JUN-2000 (Rel. 64, Created)
 XX DT 20-JAN-2001 (Rel. 66, Last updated, Version 5)
 DE Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone RP11-433K2
 KW HTG; HTGS_DRAFT; HTGS_PHASE1.
 XX OS Homo sapiens (human)
 XX OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
 XX OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RA Plumb B.;
 RT ;
 RL Submitted (20-JAN-2001) to the EMBL/GenBank/DBJ databases.


```
FT /note="assembly_fragment:00356"
FT 148878..151529
FT /note="assembly_fragment:00391"
FT 151630..155541
FT /note="assembly_fragment:00444"
FT 155642..159186
FT /note="assembly_fragment:01041"
FT 159287..162766
FT /note="assembly_fragment:01610"
FT 162867..166798
FT /note="assembly_fragment:01755"
XX
SQ Sequence 166798 BP; 45603 A; 37045 C; 37398 G; 43934 T; 2818 other;

Query Match      98.1%; Score 1093.2; DB 30; Length 166798;
Best Local Similarity 99.5%; Pred. No. 1.2e-272;
Matches 1108; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY 1 attccagatcatgagatgttgaattggtattattgagaacatgcagaaattatgct 60
Db 154134 ATTCCAGATCATGAGATGTTGAAATGGTTATTATTGAGAACATGCGCAGAAATTATGCT 154075

QY 61 attctcattagatctctgttctttctccacagatccttgttttaatttccctctaa 120
Db 154074 ATTCTCATTAGATCTCTTGCTTTCTCCACAGATATCCTTTGCTTTAAATTTCTCTTAA 154015

QY 121 gatgataaactccctgggtttattaccatacaaaactctctttatccacaagcagctt 180
Db 154014 GATGATCAAACTTCCTGGTGTATTATACCATATAAACTCTCTTTATCCACAAGCCAGCTT 153955

QY 181 tgaatttcagcaaacaccatcctctctttctccacatcttcacctttgttttcagtea 240
Db 153954 TGGAAATTCAGCAAACACCATCCTTCTTTTCCACATCTTCACCTTGTTCAGTCAC 153895

QY 241 caggttaagtcatgacatgataattagtcacctgtctctccatccacatactgtgct 300
Db 153894 CAGGTTAAGTCCATTGACATGATAATTATGTCACCTGTCTCTCATCCACATCTGCTGCT 153835

QY 301 ctccactcaggcaaatattggttccttagactctttgtttgttcacagaatactcaggatga 360
Db 153834 CTTCCACCGAGGCANATYTTGGTGCTTAGACTTCTTTGGTTTCACAGAACTACAGGATCA 153775

QY 361 tcttaggtataaggtcattgtctttttaacaaggtgtaggggacctcccatctgcac 420
Db 153774 TCTTAGGTATAAGGTCAATGCTCTTTTAAACAAGGTGATGAGGGGCTCTCCATCTGCAC 153715

QY 421 ccctgtccctcctgagtgctctccaggccatcatcagccccagcatcttctccttggaac 480
Db 153714 CCCTGCCCTCCTGAGTGTGCTCCAGGCCATCATCAGCCCCAGCATCTTCTCTTGGCAAA 153655

QY 481 gctcaaacatcctcttgcgaagtcacatcttaggattcttctcttctctcctggtctctaa 540
Db 153654 GCTCAAAACATCCTCTGCAAGTCACATCTTAGAGATCTTCTCTTCTCATGGTGGTCTCAA 153595

QY 541 catgttcattggtgtaattctctgtgtacactgcggtaccctccagtgaaacggggcca 600
Db 153594 CATGTTCAFTTGGTGAATCTTCTGTGTACACTGCGGCTACCCCCAGTGAACGGGGCCA 153535

QY 601 gctctctgtgtgtacagacacttctctcttttgcctatgactacacccacagagagac 660
Db 153534 GTCCTCTGTTGTACACAGCACTGTCTCTTTTGGCCCATGAGCTACACCCACAGGAGAC 153475

QY 661 tgtttttcacactaatgactttgagggatgctcacctttatagggttcatggtcctctcaa 720
Db 153474 TGTTTTTACACTATGACTTTGAGGGATGTCACCTTATTAGGGTTTCATGGTCTCTCAA 153415

QY 721 gaggtctacatggtgattattttatcacagacaataagaggtctatctcagtgcttccacga 780
Db 153414 GAGGCTACATGGTGATTTATTTATACACACAATAGAGGCTATCTCAGTGGCTTCACACA 153355

QY 781 gccagcctgtcccgagctccaccagtgtaaaagagcctcccgaggtatcttactgctggt 840
|||||
```

```
Db 153354 GCCAGCCTGTCCCGAGTCTCACCAGTGAAAGAGAGCCTCCCGAGGTATCTTACTGCTGGT 153295
QY 841 gagtttcttccatctactcgggtggagctttacgtttctctcatttccagaggtgtgacatg 900
Db 153294 GAGTTTGTCTTCACATACATCGGTGGACTTTTACGTTCTCATTTTCAGAGGTGTGACATG 153235
QY 901 gataaatgattctctgctagtgtggtccagggttattgtggccaatagctatgcgcgaat 960
Db 153234 GATAAATGATTCTCTGCTAGTGTGCTCCAGGTTATTGTGCGCAATAGTATGCGCAAT 153175
QY 961 tagtctttgatgctaatttatgctgataaccaaataattcaagactctgcaaatgttatg 1020
Db 153174 TAGTCTTTGATGCTAATATTATGCTGATAACCAAAATATTCAAGACTCTGCAATGTTATG 153115
QY 1021 gtttaaatattgtctctcctccaaagctcatgttgaaatttaacgcgaatgtggcagtac 1080
Db 153114 GTTTAAATATTGTCTCTCTCCAAAGCTCATGTTGAATTTAATGCCCAATGTGCGCAGTAC 153055
QY 1081 taagaagtgatgtagagaggttaattccattcatg 1114
Db 153054 TAAGAAG---TGATGAGAGGTTAATCCATTTCATG 153024

RESULT 3
AC104335/C
LOCUS
DEFINITION
6 unordered pieces.
AC104335 AL359674
AC104335.1 GI:17426299
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 217087)
Kaul, R. K., Olson, M. V., Raymond, C. and Haugen, E. D.
Direct Submission
Unpublished
2 (bases 1 to 217087)
Kaul, R. K., Olson, M. V., Raymond, C. and Haugen, E. D.
Direct Submission
Submitted (08-DEC-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Dec 8, 2001 this sequence version replaced gi:9801069.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: SC
----- Project Information
Center project name: chr-1
Center clone name: RP11-433K2 (sc0158)
----- Summary Statistics
Sequencing vector: plasmid; 50% of reads
Sequencing vector: plasmid; 108752; 50% of reads
Chemistry: Dye-terminator ET; 64% of reads
Chemistry: Dye-terminator Big Dye; 36% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 213380 bases at least Q40
Consensus quality: 215084 bases at least Q30
Consensus quality: 216045 bases at least Q20
Insert size: 256066; 17.2% error; agarose-fp
Insert size: 216587; sum-of-contigs
Quality coverage: 6.3x in Q20 bases; agarose-fp
Quality coverage: 7.5x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
```

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 4074: contig of 4074 bp in length
 * 4075 4174: gap of unknown length
 * 4175 9708: contig of 5534 bp in length
 * 9709 9808: gap of unknown length
 * 9809 13642: contig of 3834 bp in length
 * 13643 13742: gap of unknown length
 * 13743 52846: contig of 39104 bp in length
 * 52847 52946: gap of unknown length
 * 52947 140430: contig of 87484 bp in length
 * 140431 140530: gap of unknown length
 * 140531 217087: contig of 76557 bp in length.

FEATURES

source

1. .217087
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-433K2"
 /library="human BAC library 11"
 1. .4074
 /note="assembly_name:Contig168"
 4175. 9708
 /note="assembly_name:Contig169"
 9809. 13642
 /note="assembly_name:Contig170"
 13743. 52846
 /note="assembly_name:Contig171"
 52947. 140430
 /note="assembly_name:Contig172"
 140531. 217087
 /note="assembly_name:Contig173"
 60435 a 48981 c 48285 g 58724 t 662 others

BASE COUNT 60435 a 48981 c 48285 g 58724 t 662 others
 ORIGIN

Query Match 98.18; Score:1093.2; DB 2; Length 217087;
 Best Local Similarity 99.58; Pred. No. 1.2e-272;
 Matches 1108; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

Qy 1 attccagatcagagatgttgaattggtattattgagacatgagcagaaattatgct 60
 Db 121154 ATTCAGATCATAGAGATGTTGAATTGGTATTATTAGACACATGGCAGAAATATGCT 121095
 Qy 61 attctcattagatctctgctttttccacagatatacctttgctttaattttccttctaa 120
 Db 121094 ATTCATATTAGATCTTGTCTTTCTCCACAGATATCCTTTGCTTTAATTTTCTCTTAA 121035
 Qy 121 gatgataaacttccgtgtttttattaccatacaaaattctttttaccagaagccagctt 180
 Db 121034 GATGATCAAACTTCTGTGTTTTATTACCATATAAATCTTCTTTTATCCACAAGCCAGCTT 120975
 Qy 181 lggaaattcagcaaacacatctctcttctttccacatcttccaccttgttttcagtcga 240
 Db 120974 TGGAAATTTCCGAACACACATCCTTCTTCTTTTCCACATCTTCACATCTTGTGTTTCAGTCA 120915
 Qy 241 caggctcaagtcattgacatgataattagtcacactgctctctctcaccacatactgctgct 300
 Db 120914 CAGGTCCTAAGTCCATTGACATGATAAATTAGTCACCTGTCTCTCATCCACATCTGCTGCT 120855
 Qy 301 cttaactcaggcaaatatgggtgctccttagacttcttttggttcacagaatactcaggatga 360
 Db 120854 CTTTACCAGCAAAATPATTGGTGTCTTGTAGACTTCTTTGTTTCCACAGAAATCTCAGGATGA 120795
 Qy 361 tcttaggtataaggtcattgtctttttttaaacaaggtgatgaggggacctctccatctgcac 420
 Db 120794 TCTTAGGTATAAGGTCATTGTCTTTTAAACACAGGTGATGAGGGGCCCTCTCCATCTGCAC 120735
 Qy 421 cccctgctcctctgagctgtgctccaggccatcatcagccccagcagcatcttctccttgga 480
 Db 120734 CCCCTGCTCCTGAGTGTGCTTCCAGGCCATCATCATGCCCCAGCATCTTCTCTCTTGGCAAA 120675

Qy 481 gctcaaacatctctctgcaagtcacatcttaggattcttcttttctcaggtcctcaaa 540
 Db 120674 GCTCAAAACATCTTCTGCAAGTCACATCTTAGGATPCTTCCTTTTCTCATGGGTCTCTCAA 120615
 Qy 541 catgttcattggtgttaactctctgctgtacactgcggtacaccccccagtgaaacggggcca 600
 Db 120614 CATGTTTCATTGGTGTAAATCTTCTGCTGTACACTGGCGGTACCCCCAGTGAAACGGGGCCA 120555
 Qy 601 gtctctctgtttgttcatacagcaactgttctcttttttcccatgagctacacccacagagagac 660
 Db 120554 GTCTTCTCTGTTTGTATACAGCAGCTGTTCTCTTTTGTCCCATGAGCTTACACCCACAGAGAC 120495
 Qy 661 tgtttttcacactaatgactttgagggatgacacccctttataggggttcattggtccctctcaa 720
 Db 120494 TGTTTTTTCCACACTAATGACTTTTGGAGGATGTCACCTTTATAGGGTTTCATGGTCTCTCAA 120435
 Qy 721 gaggtcatatggtgattatttttatcacagacaataagaggtcatctcagtgcccttcacgca 780
 Db 120434 GAGGCTACATGGTGATTATTTTATACACACAATAGAGGCTATCTCAGTGCTTCCACACA 120375
 Qy 781 gccagctgtccccgaggtctcaccagtgtaaaagagcctccaggctatcttactgctggt 840
 Db 120374 GCCAGCCTGTCCCCGAGTCTCACCAGTGAAGAGAGCCTCCAGAGGCTATCTTACTGCTGT 120315
 Qy 841 gagttttgttcacatactagggtggaactttacgtttctcatttttcaggaggtgtgacatg 900
 Db 120314 GAGTTTTGTCTTCACATACACTGGGTGGACTTTTACGTCTCATTTTCAGAGGTTGACATG 120255
 Qy 901 gataaatgatctctgtcagtgctccaggttatgtggccaatagctatgcccgaat 960
 Db 120254 GATAAATGATCTCTGCTAGTGTGGCTCCAGGTTATGTTGGCCATAGCTATGCGGCAAT 120195
 Qy 961 tagtctttgatgctaatattatgctgatacaaaaatttccaagactctgcaaatgttatg 1020
 Db 120194 TAGTCTTTGATGCTTAATTTATGCTGTATAACCAAAATATTCAAGACTCTGCAAAATGTTATG 120135
 Qy 1021 gtttaaatattgtctctcccaagctcatgttgaatttaattcccaatgagcagctac 1080
 Db 120134 GTTTAAATATTGTCTCTCTCCAAAGCTCATGTTGAAATTTAATGCCAATGTGGCAGTAC 120075
 Qy 1081 taagaagtgatgatgagaggttaattccattcatg 1114
 Db 120074 TAAGAAG--TGATCAGAGGTTAATCCATTTCATG 120044

RESULT 4

AL390728/c
 LOCUS AL390728 206832 bp DNA linear HTG 06-FEB-2002
 DEFINITION Homo sapiens chromosome 1 clone RP11-488L18, *** SEQUENCING IN
 PROGRESS ***, 2 unordered pieces.
 ACCESSION AL390728
 VERSION AL390728.22 GI:18614040
 KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (sites)
 Direct Submission
 Submitted (05-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Feb 7, 2002 this sequence version replaced gi:18491329.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BA488L18
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Et-amersham; 9% of reads Chemistry:
 Dye-terminator Big Dye; 90% of reads
 Consensus quality: 202499 bases at least Q40
 Consensus quality: 203211 bases at least Q30
 Consensus quality: 203559 bases at least Q20
 Insert size: 206732; sum-of-contigs
 Insert size: 205638; 4.3% error; agarose-fp
 Quality coverage: 9.9ix in Q20 bases; sum-of-contigs Quality
 coverage: 10.88x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

* 1 204689: contig of 204689 bp in length
 * 204690 204789: gap of 100 bp
 * 204790 206832: contig of 2043 bp in length.

FEATURES

source
 Location/Qualifiers
 1..206832
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-488L18"
 /clone_lib="RPC1-11.2"
 misc_feature
 1..204689
 /note="assembly_fragment:03301"
 204790..206832
 /note="assembly_fragment:03801"
 58135 a 44577 c 44388 g 59631 t 101 others

Query Match 48.6%; Score 541; DB 2; Length 206832;
 Best Local Similarity 77.7%; Pred. No. 2e-129;
 Matches 813; Conservative 0; Mismatches 205; Indels 29; Gaps 12;
 Qy 1 attccagatcatagaga----tctgaaattgtgtattattgagaacatggcagaaa--- 53
 Db 126599 ATTCCAGATCAGACAGATTTATTGACATATATGATCTAATTAACACATATGGCAGAAATGT 126540
 Qy 54 ----tatgtctatctcattagatctctgtctttctccacagatatcccttcttctaatt 110
 Db 126539 TCCTAATGCAATATCATTTGATGCTTATTATTTCTCTACAGATATCCTTTATTTGTTT 126480
 Qy 111 ttccttctaag--atgatcaaaactctctgtgtttattaccatacaaatctctttatc 167
 Db 126479 AATTTTCCATGGAATACCAAGCTTTTCAGCTATATTGTCTATTAATAAATGTTTATTACC 126420
 Qy 168 cacaagccagctttggaatttcagcaaacacacatctctctctttttccacatcttcaact 227
 Db 126419 CTCAGTCAGCTTTGGATCTCAGCAACACCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 126360
 Qy 228 ttgttttcagtcacaggtctcaagtcacatgataatttagtcaacctgtctctcattcc 287
 Db 126359 TTGCTTTATCTACACAGGCTTAAGCCATTGACATGACCATTTAGTCACCTGCCCCTAATCC 126300
 Qy 298 acatactgctctcttcaactcaggaatattggtgtccttagactcttctgttttcacaga 347
 Db 126299 ACATCTGTCTACTTCTACTCAGCGGCTACTGCTGCTCCGAGCTTATTGGAATCATGGA 126240
 Qy 348 atactcag---gatgtcttaggtataggtcattgtcttttttaacaaggtgatgagg 404
 Db 126239 ATATTGACAAATGATCTAAGTGTGAAGATCATCATATTTTAAACAGGCTGATGAGG 126180
 Qy 405 gctctccatctgcacccctctcctctgagtggtgtgtccagggccat---catcagcccca 461
 Db 126179 GAGTCTCCATCTGCAACCATTTGCTCTTAAGCGTGTCTCCAGGCCATCACCATCATGACCCCA 126120

Qy 462 gcattctctcttggaagctcaaacatctcttgcgaagtcacatcttagattcttcc 521
 Db 126119 GCACCTCTCTTGGAAAAGTTTAAATATTTCTGCAATACACACCTTAGCTTCATCC 126060
 Qy 522 tttctcatgggtctcacaatgttcattgtgtg--taactctctctgtacatcggccta 580
 Db 126059 TCTTCTCATGGGTCTCAACATGTTCAATTAATACCTTTTCTTTCATTTGCTTA 126000
 Qy 581 ccccaagt-gaaaagggggccagctctctgtttgtctacacacacgttctcttttggcca 639
 Db 125999 CCCGGAATAGATTGGGGCCAGTCTTTTGTGTCACCTGAGCACTGTATGTTGGCCA 125940
 Qy 640 tgaactacacacacagagagactgttttttcacataatgaacttgaaggatgcacctta 699
 Db 125939 TGAGCTACACCCACAGAGGCTGTTTTCATCTAATATGTTTGGAGGATGTCATCTCA 125880
 Qy 700 taggggttcaggtctcctcacaagaggtcacatggtgattattttatcacagacaagaagg 759
 Db 125879 TAGGACTCATGGTCCCTCTCAAGTGGCTA-TGGGTGATATTATTTGTACAAACA-AAGAGA 125822
 Qy 760 ctatctcagtccttcacagcagcagcctgtctccc--gagctcaccagtgaaaagagcc 817
 Db 125821 CTGTCCAGCAGCTACATAGCACCAGCGTCTCCCAAGAGCCTCCCAGCAAAAAGGGCC 125762
 Qy 818 tccaggtctattctactgtgtgtga---gtttgtctcactactgggtgagactttacg 874
 Db 125761 TCCAGACCATCTTATTTGTTGGTGGAGCTGTTTGTCTTATATATTTGGTGGATTCATA 125702
 Qy 875 ttctcatcttcagaggtgtgacatgataaattgattctctgctagtggtctcaggtt 934
 Db 125701 TTCTCATTTTCAGTGGTGTGACATGATGAACGACCCCTCTACTGTATGATTCATATG 125642
 Qy 935 attgtggccaatagctatccgcaattagtcctttttagctcaatttctgctgataaccac 994
 Db 125641 CCTGTGTCATAAATCACTATGCCATAATTAGTCTTTAGCTTTACCTATGCTGTATAA-CAA 125583
 Qy 995 atattcagaactctcgaatgttatgg 1021
 Db 125582 ATATTCAAGATTGTTCAAGACTCTTTGG 125556

RESULT 5

AC104335
 LOCUS
 DEFINITION
 Homo sapiens chromosome 1 clone RP11-433K2, WORKING DRAFT SEQUENCE,
 6 unordered pieces.
 AC104335.AL359674
 VERSION
 AC104335.1 GI:17426299
 KEYWORDS
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 SOURCE
 human
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 217087)
 Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
 Direct Submission
 TITLE
 Unpublished
 JOURNAL
 2 (bases 1 to 217087)
 Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
 Direct Submission
 AUTHORS
 Submitted (08-DEC-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 JOURNAL
 On Dec 8, 2001 this sequence version replaced gi:9801069.
 COMMENT

 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: http://www.genome.washington.edu
 Contact: uwgchgs@u.washington.edu
 Drafting Center: SC

 Project Information
 Center project name: chr-1
 Center clone name: RP11-433K2 (sc0158)

 Summary Statistics

estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

FEATURES	source	Location/Qualifiers
		1. '73661
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="16"
		/clone="CTD-2358C21"
BASE COUNT	20517 a 16348 c 15807 g 20989 t	
ORIGIN		

Query Match	28.08;	Score 311.6;	DB 9;	Length 73661;
Best Local Similarity	62.7%;	Pred. NO. 5.6e-70;		
Matches 638;	Conservative 0;	Mismatches 359;	Indels 21;	Gaps 9;
QY	15	agatgtgaaatgggtattattagagacaatgagcagaataattatgctattctcattagatc 74		
Db	73636	AAAGATGATGATGGTAAACAATTTTGAACACTGAACAAACCAACATTTGATAACATGGGATC 73577		
QY	75	tcttgctt-----ttctcacagatatacctttgctttaaattttcccttctaagaatgatca 128		
Db	73576	TATTGTTTCTTTTATTCAGATATAGAAAGTATGTTTCCCTCTATGCCATTAAAGATGAAC 73517		
QY	129	aacttcctggttttattaacatacaaaattctttttatcacagaacagcagcttggaaattt 188		
Db	73516	TATTTCTACTTATGCTATCGTTTAAAAATGCCTTGTGTTTTCAGCTGGCATTTGGATTTCT 73457		
QY	189	cagcaaacacacatcctctctcttttccacatcttcaaccttgttttcagtcacaggtcta 248		
Db	73456	CAGCAACACCTTCTCCTTTTCTTCCACATCTTTAGGCTTCTTCTGAAATCGCAGGCGTA 73397		
QY	249	agtcattgacatgataattagtagtcacctgtctctctccacatactctgctcttcaactc 308		
Db	73396	AAOCCCGTGACTTTACTACACCTGCACCTGGCCCTCATTTACATTCAGATGCTCTCACTG 73337		
QY	309	aggcaaatattggctgcttagaactcttttgggttcacagaatactcaggatgactctaggt 368		
Db	73336	CAGTGGATTTTTGGCTCTAGACATATTTGAATACATGCAATTTGGGAATGACTTTTAAGT 73277		
QY	369	ataagtgtaattgtctttttaacaaaggatgagggggcctctccatctgacccctggcc 428		
Db	73276	GTAAGGCACTTTTTTACAAACAGGGCAATGAAGGGCCTCTCCATCTGCACCACTGCC 73217		
QY	429	tctgagtggtctccaggccat---catcagccccagcatctctctctctctctcagaaagctca 485		
Db	73216	TTCTGAACATGCTCCAGGGCCATCAGATCAGCCCCACGACCTCTGCTGGTTGGCAAGGTTTA 73157		
QY	486	aacatcctctcgaaatgcacatcttaggattcttctcttttctctatgggtcctcaaatgt 545		
Db	73156	AACATAAATCCACAATTTACATTTTGCATGTTTTCTTCTTTTGGGTGCTGCTCAATTTGT 73097		
QY	546	taattgggt-gtaattctctgctgacactggcgtacccccagtgaaacggggg-ccaagtc 603		
Db	73096	CCTTCAGTAGTTGGCCAAATCTTTTCTCTGTCATCTTCCACACGACCCCATATCAAGC 73037		
QY	604	tictgtttctacacagcactgtctctttttgccaatgagctacacccccacagagactgt 663		
Db	73036	TACAGAAATGTCAATAATCTGCCACTTTTC-CCCCATGAACCTCCACAATCAGGGGAGTGT 72978		
QY	664	tttccacactaatgactttgagggatgtcacctttatagggtttcaaggctctctcaagag 723		
Db	72977	TUUTTCACGTGACATTTATCCAAGTAGTCTCCTGTGTAGTACTTATGTCTCTCAAGGG 72918		
QY	724	gtacaaatgggtattattttatacagacaataagaggctatctcaagtccttcacgcagcc 783		
Db	72917	CATACATGCTGATCTCTCGCCACGAGC-ATTGGAGGCAATCCAGCACCTCCATAGCAC 72859		

7
RESULT
AC106730
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1
2
3
4
5
6
7

FEATURES

source

BASE COUNT

ORIGIN

Qy	784	agcctgtccccc--gagtcacacagtgaaagagcctccocaggtcatcttactgtctggtg	841
Db	72858	CACCTCTCCCCAAGACCCCTCCCCAGAGAAAAGGCCACCCAGACCATCTGCTGCTGGT	72799
Qy	842	agt---tttgctctcacatactgggtgagactttacgtcttcatttccaggaggtgtgaca	898
Db	72798	AGTTTCTTTTGGTCTATGTACTGGGTGGACCTTTATCTTCAACCTCTTTAACCCCTGCTA	72739
Qy	899	tggataaatgattctc---tgctagttggctccaggttatgttggccaatagctatgcc	955
Db	72738	TGGCCGATATGCCACACAGTTTGCTTCCCTGAGTGTGCAGAGAGCTGTGCTCAATGCCATATGCC	72679
Qy	956	gcaattagtcctttgatgctaaatttatgctgataaaccaaatattccaagactctgcaaa	1013
Db	72678	ACTCTTAGTCCTTTGGTGCAATCAGTCTGTATAAAAACAATAATCAGTATTATATGCAAA	72621

AC106730 152748 bp DNA linear PRI 03-FEB-2002
Homo sapiens chromosome 16 clone CTD-2014E2, complete sequence.
AC106730
HTG.
AC106730.2 GI:18482002
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152748)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 152748)
DOE Joint Genome Institute.
Direct Submission
Submitted (12-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 152748)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 3, 2002 this sequence version replaced gi:1819280

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

```
-----
Location/Qualifiers
1. 152748
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="CID-2014E2"
46721 a 30969 c 30999 g 44059 t
```



```
Query Match      28.0%; Score 311.6; DB 9; Length 152748;
Best Local Similarity 62.7%; Pred. No. 6e-70;
Matches 638; Conservative 0; Mismatches 359; Indels 21; Gaps 9;

QY 15 agatattgaattgattattgagacatgcgagaaattatgctattctcattagatc 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132487 AAAGATGATGATGGTAACAAATTTTGAACCTCTGACACCAACCAATGATAACATGGGATC 132546

QY 75 tcttgctt-----tctccacagatatcttctgtgttaatttcccttccatgaatgatca 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132547 TATGTTTTCTTTTATCCAGATATAGAAAGTATGTTCTCTCTATGCCATTAAGATGAAC 132606

QY 129 acttctcgtgtttattaccatcacaaattctttttatccacagcagcgtttggaattt 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132607 TATTTTCTACTTATGCTATCGTTAAATAATGCGCTTGTGTTTTCGAAGCTGCGATTTGGATTCT 132666

QY 189 cagcaaaacaccatctctcttctccacatcttcacctttgttttttccagtcacaggtcta 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132667 CAGCCNACACCTTCTCTCTTTCTTCCACATCTTTTACGGTCTCTTGAATCGCAGGCCTTA 132726

QY 249 agtccattgacatgataattagtcacctgtctctcatcacatcactgctcctcactc 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132727 AACCCCGTGACTTACTCACTCTGACCTGGCCCTCATTCACATTCAGATGCTCCTCACATG 132786

QY 309 aggcataattggtccttagactctctcttcttctcagagaataactcaggatgatcttaggt 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132787 CAGTGGATTTTGTGCTCTGACATATTTGAATCATCTGCAATTTTGGGAATGACTTTTAAGT 132846

QY 369 ataaggtcatctctctttaaacaaggtgatgaggggcctctccatctcacccccctgpc 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132847 GTAAGGCATTTTATACACAAACAGGGCAATGAGGGCTCTCCATCTGCACCACTGCTGCC 132906

QY 429 tcttgagtgctccaggccat---catcagccccagcactctctcctctggcacaagctca 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132907 TTCTGAACATGCTCCAGGCGCATCAGCATCAGCCCGCCAGCACCTCTCTGTTGGCAAGTTTA 132966

QY 486 acatctctcagtcacatcttaggattctctcttctctcagtggtccctcaacatgt 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132967 ACATATAATCCAAATTAATTTGTCATGTTTCTCTTTTGGGTGCTCTCAATTTGT 133026

QY 546 tcatgtgtt-gtaattctctgtgtacatcggtcggtaccocccagtggaacggggg-ccagtc 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133027 CCTTCAGTAGTTGGCCAAATCTTTTCTCTGTGCTATCTTCCACAGCAGCCCATATCAAGC 133086

QY 604 tctgtttgtcctacagcactgttctcttcttggccatgagtcacacccacagagagactgt 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133087 TACAGAATGTCAGTAATAATCTGCCACTTTC-CCCCATGAATCCCAATCAGGGGAGTGT 133145

QY 664 ttttcacataatgactttgaggatgtcaccttttatagggttcatggttctcctcacaagag 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133146 TTTTCTCTGTGACATTATCCAAAGTATGCTCTCTGTGTAAGTATTATGCTCTCAAGGG 133205

QY 724 gctacagtggtattttttatcacagaaataagagctatctcagtgctcttcacgcagcc 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133206 CATACATGCTGATTTCTCTGCCAGGC-ATTGGAGGCAATCCAGCACCTCCATAGCACCC 133264

QY 784 agcctgtcccc--gagtcctcacagtgaaagagcctccagggctatcttactgtcgtgtg 841
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133265 CACCTTCCCAAGACCCCTCCCAAGAAAGGGCCACCCAGACCATCTGCTGTGCTGTG 133324

QY 842 agt---tttgctctcacactcagtggtgagcttttagcttctcatttcagagggtgtgaca 898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133325 AGTTCTTTGTGGTGATGACTTGGGTGGAGCTTTATCTTCTCAACCTCTTTTACCCCTGCTTA 133384

QY 899 tggataaatgattctc---tgctagtgtggtccaggtttatttggcccaatagctatgcc 955
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133385 TGGCCGTATGACCCACACCTGTGCTCTGAGTGTGCAGAGAGGCTGTGCTCAATGCCATGCC 133444

QY 956 gcaattagtcctttagtgtaatttatgctgataacccaataattcaagactctgcataa 1013
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133445 ACTGTTAGTCTCTTTGGTGCAAAATCAGTTCTGTGATAAAACAATAATCATGATTATGCAAA 133502
```

```
RESULT 8
AC073188/c
LOCUS
DEFINITION
AC073188
ACCESSION
AC073188.10 GI:15920117
VERSION
HTG..
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 65128)
Substont,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
2 (bases 1 to 65128)
Cordum,H. and Abbott,A.
The sequence of Homo sapiens BAC clone RP11-73B2
Unpublished (2001)
3 (bases 1 to 65128)
Waterston,R.H.
Direct Submission
Submitted (10-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 65128)
Waterston,R.H.
Direct Submission
Submitted (04-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 65128)
Waterston,R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 4, 2001 this sequence version replaced gl:14647308.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
----- Center project name: H_NH0073B02
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7>, send <mailto:regreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male

donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-905H7, 2000 bp overlap; the clone sequenced to the right is RP5-1193P9, 2000 bp overlap. Actual start of this clone is at base position 36351 of RP11-73B2; actual end is at base position 72100 of RP5-1193P9.

Polymorphisms exist between AC073188 and AC006457. A single plasmid region exists between 68464-68496 and 68531-68559. A single stranded/chemistry region exists at base 68546. This clone contains an extra sequence fragment not represented in the finished neighbor, AC006457. This is submitted as H_NH0073B02_F1.

FEATURES

```
source
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="7"
    /map="7"
    /clone="RP11-73B2"
    /clone_lib="RPC1-11"
    406..1310
    /rpt_family="L1"
    1317..1547
    /rpt_family="L1"
    3257..3634
    /rpt_family="L1"
    3635..3925
    /rpt_family="MaLR"
    3930..4120
    /rpt_family="MaLR"
    4136..4334
    /rpt_family="Alu"
    4341..4488
    /rpt_family="Alu"
    4555..4753
    /rpt_family="L1"
    5597..5693
    /rpt_family="MER1_type"
    5802..6205
    /rpt_family="L1"
    9188..9281
    /rpt_family="MER53"
    10528..10758
    /note="similar to EST AI028332 (NID:g3245641) ov90g03.xl"
    10576..10799
    /rpt_family="MaLR"
    10813..11168
    /rpt_family="MaLR"
    11452..11647
    /rpt_family="MER2_type"
    11645..11696
    /rpt_family="TATGn"
    11701..11833
    /rpt_family="Alu"
    12521..13095
    /note="similar to EST AI806860 (NID:g5393426) wf36b07.xl"
    13058..13087
    /rpt_family="AT_rich"
    13146..13198
    /rpt_family="AT_rich"
    13365..13671
    /rpt_family="Alu"
    14087..14115
    /rpt_family="(TTTTTA)n"
    14129..14883
    /rpt_family="L1"
```

```
repeat_region 15223..15246
                /rpt_family="AT_rich"
misc_feature 15290..15571
                /note="similar to EST AW445087 (NID:g6986849)"
misc_feature 15290..15571
                /note="similar to EST AW627961 (NID:g7374751) hi36d12.xl"
repeat_region 16305..16574
                /rpt_family="Alu"
repeat_region 16753..17021
                /rpt_family="Alu"
repeat_region 17038..17286
                /rpt_family="Alu"
repeat_region 17700..18467
                /rpt_family="L1"
repeat_region 17853..17872
                /rpt_family="(CAAAA)n"
repeat_region 18468..18778
                /rpt_family="Alu"
repeat_region 18752..18796
                /rpt_family="(A)n"
repeat_region 18779..18938
                /rpt_family="L1"
repeat_region 18838..18870
                /rpt_family="AT_rich"
repeat_region 18944..19246
                /rpt_family="Alu"
repeat_region 19220..19265
                /rpt_family="(GAAA)n"
repeat_region 19256..19691
                /rpt_family="ERV1"
repeat_region 19748..19790
                /rpt_family="(CAAAA)n"
misc_feature 20557..20613
                /note="similar to EST AW445087 (NID:g6986849)"
misc_feature 20557..20666
                /note="similar to EST AI028332 (NID:g3245641) ov90g03.xl"
misc_feature 20594..20828
                /note="similar to EST AA292895 (NID:g1941876) zt66c08.rl"
repeat_region 21125..21428
                /rpt_family="Alu"
repeat_region 21664..21691
                /rpt_family="AT_rich"
misc_feature 22442..22750
                /note="similar to EST BF899282 (NID:g12290741)"
misc_feature 22640..23050
                /note="similar to EST AI279713 (NID:g3917947) ql53f03.xl"
misc_feature 22855..23086
                /note="similar to EST AA292895 (NID:g1941876) zt66c08.rl"
repeat_region 23402..23451
                /rpt_family="T-rich"
repeat_region 23430..23733
                /rpt_family="Alu"
repeat_region 23850..24148
                /rpt_family="MER2_type"
repeat_region 24565..25150
                /rpt_family="L1"
repeat_region 25151..25275
```

Query Match 27.2%; Score 303; DB 9; Length 65128;
Best Local Similarity 62.0%; Pred. No. 9.5e-68;
Matches 638; Conservative 0; Mismatches 365; Indels 26; Gaps 9;

```
QY 11 ataggagattgaaattggtattattgagaaatggcagaattatgtattctcat 70
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37223 ATTAAATGATAATATTTGACCGTGAATAATATTAATAATATGATGATAGGA 37164

QY 71 gatctctgcttttccacagatatccttgcctttaatttccttcttaagatgacaa 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37163 TATCTATTGCTTTTCAGATATATAAATATGATCCCTCTCCACCATTAAGACGCNTA 37104

QY 131 cttcctgggttttattaccatacaactctttttatcccaagccagcttggaattca 190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37103 G-TCCACTTATGACACTATTAAATATGCCCCCTTTTCCCAAGCTGTTATGGACTACA 37045
```


QY 371 aaggtcattgtctttttaaacbaagtgatgagggccctctccatctgcaaccccccgcctc 430
DB 118279 AAGGTATTCTTACATGCATGAGGGATGAGGG--CTCTCCATCTGCACCACTGCCTC 118336
QY 431 ctgagtgctgcaccagg---catcatagcccccagcatctctcttgccaagctcaaa 487
DB 118337 CGGAGTGTGCTCCAGCGTGTACCAATATACCCCGGACGCTCTGTTGGCAAGAAATAAA 118396
QY 488 catctcttgcaagtcacatcttaggattctctctttctctctctggtctcacaatgttc 547
DB 118397 CAGAAATTCACACATTTGCATCTTCTTCTTGGGTTCTCAGTTT-----CTTTC 118447
QY 548 attggtgtaattctctgtgtacactgaggctacccccagtgaaacggg-gccagttcttc 606
DB 118448 TCAGTAGTAACCTGCTTCTCTCCACTGTGGCTCTTCTTAATGTGACCGACACAAATGTGC 118507
QY 607 tttgtgtacacagcactgtctcttttttgcccatgactgtcacccacaggaagactgttt 666
DB 118508 CAAGTATCAGTAATAACTGTTCACCTTCTTACATAAGCTACATCATCAGGAGTCTGTCTT 118567
QY 667 teacactaatgactttgagggatgtcaacttttataggggtctctctctcaagaggtc 726
DB 118568 TCATGCTGCCACTACTTGCAGATGCTCTTCTTGTAGCAGTCACTGTGCTCTCAAGGGCAT 118627
QY 727 acatggtgattattatcacagacaataagagctatctcagtgctctcagcagccagc 786
DB 118628 ACATGGTAATCTTTTGTGCCAGC-ATCAGACACACCCAGTACTTCCCAAGCACCAAC 118686
QY 787 ctgtc---ccgaggtctcaacagtgaaaagagcctccagggtactcttactgctggtgagt 844
DB 118687 CTGTCTGCAAAAGCTCCCGCAGAGAAAGGGCCACCTAGACCATTCCGATGTGTTGAGT 118746
QY 845 ---ttgtcttcacactaggtgagctttacgttctcaatttcacaggaggtgtgacatgg 901
DB 118747 TCGTTGTGCTCATATATGGGTGGGCTCTCATCTCATCTCTCTCAACCCCTGTATGG 118806
QY 902 atcaatgattctctgtagtggtccaggttatgtggccaatagctatccgcaatt 961
DB 118807 GCTTATGACTCATTCATCATGCGGTGTCAGATGCTTATGGGCAATTTCTAT---CTGTT 118862
QY 962 agtctttgattgtaattatgctgataccaaattattcaagactctgcaaatgttatgg 1021
DB 118863 AGCCCTTTTATGCTAATATAGTTCTGATATAAGATAAATAATCAAAATTTTGCAATATGTTA 118922
QY 1022 tttaaatat 1030
DB 118923 TATAAATAT 118931

RESULT 10
AB064662
LOCUS
DEFINITION
AB064662
Capra hircus gene for putative pheromone receptor gVIR1, complete cds.
VERSION
AB064662.1 GI:18148935
KEYWORDS
Capra hircus (strain:Shiba) DNA.
SOURCE
Capra hircus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Caprinae; Capra.
REFERENCE
1 (sites)
Wakabayashi, Y., Mori, Y., Ichikawa, M., Yazaki, K. and Hagino-Yamagishi, K.
AUTHORS
Gene of goat putative pheromone receptor is expressed in two distinct olfactory organs
TITLE
Unpublished
JOURNAL
2 (bases 1 to 989)
REFERENCE
Wakabayashi, Y., Mori, Y., Ichikawa, M., Yazaki, K. and Hagino-Yamagishi, K.
AUTHORS
Direct Submission
TITLE

JOURNAL Submitted (03-JUL-2001) Kimiko Hagino-Yamagishi, Tokyo Metropolitan Institute of Medical Science, Department of Ultrastructural Research; 3-18-22 Honkomagome, Bunkyo-ku, Tokyo 113-8613, Japan (E-mail:k.yamagis@ins.hoken.or.jp, Tel:81-3-3823-2101(ex.5345), Fax:81-3-3823-2965)

FEATURES
source
1..989
/organism="Capra hircus"
/strain="Shiba"
/db_xref="taxon:9925"
1..927
/gene="gVIR1"
1..927
/gene="gVIR1"
/note="Shiba goat putative pheromone receptor gVIR1"
/codon_start=1
/product="putative pheromone receptor gVIR1"
/protein_id="BAB83523.1"
/db_xref="GI:18148935"
/translation="MKKNRLSSSIDIRNAIFSEVAIGILANTILLFHAHNPFLLEHRP
KSTDTTIGHLALIHIVMLLTAFMATDFTGSKTWDDIQCKLVYLYLSSMRGLSICAT
CLLSVLMAITLSPRNLSLAKEKLRPSHNLYCTFLFWFNPFINGSLSTIATPNVT
SAHLRVTESCLRPVHFRLYLOFALFTIQDLCLLGLMALSSGYMTVLLYLRHKRQTQ
ELQSTRSPKASPEERATQTILFMSFFVMYFLDITVSWFSRMLWDIDSVRMCVQML
VGNGYASTPLVLISLEKRIIKLVKIWRNH"
BASE COUNT 240 a 267 c 193 g 289 t
ORIGIN

Query Match 27.1%; Score 302; DB 4; Length 989;
Best Local Similarity 63.2%; Pred. No. 1.2e-67;
Matches 581; Conservative 0; Mismatches 325; Indels 14; Gaps 7;

QY 119 aagatgatacaactctcgtgtttttattaccatacaaaattctttttatccacaagccagc 178
DB 4 AAAAAAATAGACTTCTCCAGTTCCATTGACATGAAGAAATGCCATTTCTCCGAAGTTGC 63
QY 179 ttgtgaatttcagcaaacaccatctcttcttctccacatcttccactttgttttcaagt 238
DB 64 ATTGGATCTTGGCCCAACACCATCTCTTCTCTCCATGCCCAACACACTCTCTTTGAG 123
QY 239 cacaggtctaaagtcattgacatgataattagtcacctgtctctccatccacatactgctg 298
DB 124 CACAGACCCCAAGTCCACTGACCTGACCAATTTGGTCACTTGGCCCTAATCCACATAGTCATG 183
QY 299 ctcttcactcaggcaaatattggtgtccttagactcttcttggttcacagaataactcagat 358
DB 184 CTACTAACTGTGGC---ATTATGGCTACAGACACACTTTGGGTCTCAGAAAAACCTGGGAT 240
QY 359 gatcttaggtataagtgcatgtctttttaacaaggtgatgaggggctctccactctgc 418
DB 241 GACATCCCAATGTAAAGTTGGTTGTCTACCTGTACAGTTTCATGAGGGGCTCTCCATTGCT 300
QY 419 acccctgctcctcgtgagtggtgtccaggccatca---tcagccccagcatcttctcttg 475
DB 301 GCCACTGCTGCTGAGTGTCTCTGCGGCATCACCTTCAGCCCCAGAAACTCTCTGCTG 360
QY 476 gcaagctcaaacactctctgtcgaagtcacatttaggattcttctcttctcattgaggtc 535
DB 361 GCCAAGTTCAAACTTCGCCCATCCCATCACTGTATTGCACTCTCTTCTCTGGGTC 420
QY 536 ctcaacatgttctcagtggtgtaatacttc-tgctgacactgcggtaccctccagtg-aaac 593
DB 421 TTCAACATGTTTCATCAATGGTTCTTCTTACTGTGCACATTTGCCACCCCAACGAGTGC 480
QY 594 ggggagcagctctctgtttgtctacacagcaactgttctcttttgcctcagctacacccac 653
DB 481 TCAGCCCATCTTCTACGTGTCACTGAATCTCTCACTTCGCCCCCGCTGATTCACTTCTCT 540
QY 654 aggagactgtttttcacactaatgactttgagggatgtcacctttataggggttcattggtc 713
DB 541 AGGTATTTACAGTTTGCACACTGAGGACCAATTCAGACACATCTGCTTGTGGGGCTCATGGCC 600

```
Qy 714 ctctcaagaggtacatggtgattattttatcacagacaataagaggctatctcagtgctt 773
Db 601 CTCTCAAGTGTGATACATGCTGACTCTCTTGTACAGGC-ATAAGAGCCAGACCCAGCATCT 659
Qy 774 tcaagcagcagcagctgtccccc-gagtcacacagtgtaaaagagcctcccccaggtatctt 831
Db 660 TCAGAGCACCAGGCTTCTCCAAAAGCATCCCAAGAAAGAGGCGCCACCCAGACCATCT 719
Qy 832 actgctggtgagt---tttgtcttcacatactcgggtgagctttagcttctcattttcagg 888
Db 720 ATTCTCATGAGTTCCTTGTGGTCATGTCATTTTGGACATAAATGCTCTCTTGGTTC 779
Qy 889 aggttgacatggaataatgattctctgctagtggtgctccaggttattgtgggccaatag 948
Db 780 AAGAATGTTGGGATATCGACTCTGTTGCTATGTTGTCAGATGCTTGTGGGAAATGG 839
Qy 949 ctatgccgaattagtcctttgctgataattgctgatacccaaatatcgaagctct 1008
Db 840 CTATGCCAGCATAGCCCTTGTGGTGAATCAGTACTGAAAGAGAAATAATCAAGGTCTT 899
Qy 1009 gcaaatgttatggtttaat 1028
Db 900 AAATACATATGGTGAGAT 919

RESULT 11
AC108131
LOCUS Homo sapiens chromosome 16 clone RP11-337N9, DNA linear PRI 06-FEB-2002
DEFINITION Homo sapiens chromosome 16 clone RP11-337N9, complete sequence.
ACCESSION AC108131 AC028260
VERSION AC108131.2 GI:18542984
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 151941)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 6, 2002 this sequence version replaced gi:18369977.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
Location/Qualifiers
1. .151941
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-337N9"
BASE COUNT 42554 a 32907 c 31887 g 44593 t
ORIGIN
Query Match 25.1%; Score 279.6; DB 9; Length 151941;-
Best Local Similarity 60.3%; Pred. No. 1.2e-61;
Matches 567; Conservative 0; Mismatches 364; Indels 9; Gaps 6;
Qy 11 atagagagtgtaaaatgggttattattgagacaatggcagaataattatgctattctcatta 70
Db 70064 ATTAAAAATGAAGGTTGGTAATAATTTTGACTATACAATAAATCAATTAACATTCATGATGA 70123
Qy 71 gatctctgctttctccacagatat-cctttgcttttaatttctctcttaagatgataca 129
Db 70124 GGTCTATTGCTTTTTCAGGTATATAAAATATAAATTCCTCTCTTCATCATTAAGACGACCA 70183
Qy 130 acttctggtttattaccacatacaaatctcttctttatccacaagcagcttttgaaatttc 189
Db 70184 TGTTCAGTTAAGGTACTACTAAACAATTGCCCTTCTTTTCAAGCTGTTATTCAGCTCAG 70243
Qy 190 agcaaacacacatccttcttcttccacatcttccacacacacacacacacacacacacac 249
Db 70244 AGCAACAACCTCTCTCTATTTCTTCCAAATCTTCTCACTCTCTTCAGGATCACAGGCCATA 70303
Qy 250 gtccattgacatgataatagtcacacctgtctctccacacacacacacacacacacacac 309
Db 70304 GCCCACTGACCCCATCACCTGTCACTGGCCCTTGCCACTTTCAGGATGCTCTCTCATGT 70363
Qy 310 ggcaatattgggtgccttagaactctttgttcacagaatactcagatgatcttaggta 369
Db 70364 GGTCTTCTTGGCATCTCCAGACCTGTTGAGTCACCTGATTTTTCAGAAATGACTTCAAGTG 70423
Qy 370 taaggtcattgctttttaaacaagtgatgagggccctctccatctcagtcgacccctgcct 429
Db 70424 TAAGGATCTCTTCGTCATGCACAGGGTGATGAGGAGCCCTCTCCATTCGACCCAGCTCT 70483
Qy 430 cctgagtggtctccaggc---catcatcagccccagcatcttctctcttgcaaaagctcaa 486
Db 70484 CCTGAGCATGCTCCAGGCTGTGCGCCATCAGCCCTGGCACCTCTGCTCTGCAAGGATTA 70543
Qy 487 acatcctctgcaagtcaacatcttaggtattctctcttctctctctcagtcgacccctgc 545
Db 70544 ACAGAAATTCAAAAGGTTACATCTTTCACCTCTCTTCTCTCTATGGTCTCTCAGTTTGT 70603
Qy 546 tcattggtgtaattctctgctgacactcggtcagctaccccccagtgaaacgggcccagcttt 605
Db 70604 TCTCAGTAGTAACTGCTCTCTCTCCACTGTGGCTTCTTCTAATGGGACCAAGCGTTGT 70663
Qy 606 -ctgtttgtcacagcactgttctctttttggccatgagctacacccacagagactgtt 664
Db 70664 GCTAAGTAGTACATAAATACTGATCATCTTCTTCCATAAGCTACATTATCAGAGGCTGTC 70723
Qy 665 ttccacataatgacttgagggtgacacatttatagggttcattggttctctcctcaagagg 724
Db 70724 TTTTATGCTGCCACATGCTTACGAATGTCTCTTTTGTAGCAATCATCGACGCCCTCAAGTGC 70783
Qy 725 ctacatggtgattattttatcacagacaataagaggctatctcagtgcttccacagcagca 784
Db 70784 ATACATGTAATTCCTTGTTCAGGC-ATCAGAGACAATCCCAATACCTTTCATAGAACA 70842
Qy 785 gctgtccccagtgctcaccagtgaaaaagagcctccaggctattcttactctggtgagt 844
Db 70843 ACCTCTCTCCAAGAGCCCTCCCAAGAGAAAAGGGGACCA--TCTTGTCTGGTGAAGTTC 70900
Qy 845 ttgtcttcacatactgggtgagactttacgttctctcatttttcaggaggtgtgacatggata 904
Db 70901 TTTTGGTCAATGATCTAGGGGGAGCTTATTATCTCATCTCTCTCAACTCTATTATTGGCT 70960
Qy 905 aatgattctctgctagtggtgctccagggtatttgaggca 944
```



```

Db 148141 GGTCTCTTGCACTCCAGACCTGTTTGAGTCACTGATTTTCAGAACTGACTCAAGTG 148082
Qy 370 taaggtcattgtctcttttaaaacagggtgatgaggggccctccatctgcaccccccctgcct 429
Db 148081 TAAGGCATTCCTTCGATGCACAGGGGTGATGAGGAGCCTCTCCATCTGCACACCTGCCT 148022
Qy 430 cctgagtgctccaggc---catcatcagcccccagcatctctccctgcaagagctcaa 486
Db 148021 CCTGAGATGCTCCAGGCTGTCGCCATCAGCCCTGGCCACCTCTCTGGTCTGCAAGGATTA 147962
Qy 487 acatcctcttgcaagtcacatcttaggattctctctctttctctcaggggtccctccaacatgt- 545
Db 147961 ACAGAAATTCAAAGGTTACATCTTTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 147902
Qy 546 tcattggtgtaattctctctgctgacactggggtacccccagtgaaacggggccagctctt 605
Db 147901 TCTCAGTAGTAACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 147842
Qy 606 -ctgtttgtctacagcactgttctcttttggccatgagctacacccacagggagactggt 664
Db 147841 GCTAAGTATCATGTAATACTGATCATCTTCTTCTTCCATGACTACATTTATCAGGACCTGTC 147782
Qy 665 ttccacactaatgactttgagggtatgcacotcttatagggttcatggttctctctccaaggg 724
Db 147781 TTTGATGCTGCCACTGCTTACGAATGTCTTCTTTGTAGCAATCATGACGCCCTCAAGTGC 147722
Qy 725 ctacatggtgattatttatcacacataagaggtctatctcaatgcttccttcacgcagccca 784
Db 147721 ATACATGGTAATCTCTTGTTCAGGC-ATCAGAGACAAATCCCAATACCTTCATAGAACCA 147663
Qy 785 gcctgtcccccaggtctcaccagtgtaaaagagcctcccgaggctattctactgctggtgagt 844
Db 147662 ACCTCTCTCCAAAGCCTCCCCACAGAAAAGGGGACCA--TCTTGTGCTGTCAGTTC 147605
Qy 845 ttgtcttcacactactggtggtgacttctcttctcatttcttcagagtggtgacatgata 904
Db 147604 TTTCTGTGTCATGTAATCTGGGGGACCTTATTATCTCATCTCTCTCAACTCTATTATTGGCT 147545
Qy 905 aatgatctctgctgagtggtgctccagggtattgtggcca 944
Db 147544 TATGACTCAGTCATGTGAGCATGCAGAGGCTTCTGGCCA 147505

RESULT 13
AC022202/c
LOCUS
DEFINITION Homo sapiens BAC clone RP11-321E8 from 7, complete sequence.
ACCESSION AC022202
VERSION AC022202.12 GI:13992747
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE
AUTHORS Meyer,R., Abbott,A. and Hawkins,M.
TITLE The sequence of Homo sapiens BAC clone RP11-321E8
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 95097)
Waterston,R.H.
Direct Submission
Submitted (26-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
AUTHORS 4 (bases 1 to 95097)
Waterston,R.H.
Direct Submission

```

Submitted (08-MAY-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 95097)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 8, 2001 this sequence version replaced gi:13431144.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics

Center project name: H_NH0321E08

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 17 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/OTB/CHRI17 , send
mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-468B6, 200 bp overlap; the clone sequenced to the right is RP11-561N12. Actual start of this clone is at base position 61428 of RP11-468B6; actual end is at base position 95097 of RP11-321E8.

FEATURES
source
1. .95097
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="7"
/clone="RP11-321E8"
/clone_lib="RPCI-11"
repeat_region 273..564
/rpt_family="Alu"
repeat_region 1280..1352
/rpt_family="L2"
repeat_region 1371..1404
/rpt_family="MER1_type"
repeat_region 1570..1808
/rpt_family="L2"
repeat_region 1861..2172


```

Db 35980 GAATGTCAGTAAATACCTGCTCCTCCTTCCCAATGAATGTCACCATGAGGCTGTTTGT 35921
Qy 668 cacactaatgactttgagggatgcacactttataggggttcagtgctctctcaagagcta 727
Db 35920 TACTTGTGCGTATCCAGAGATGCTTCTCTGTAGGAATCATGCCGCTCTCAAGTGCCCTA 35861
Qy 728 catggtgattattttatcacagacaataagagdgcttatctcagtgcttcacgagccagcc 787
Db 35860 CATGGTGAATCTCTTGTCCAGGC-ATCAGAGGCGCTCCAGCACCTTCACAGCACTAGCT 35802
Qy 788 tgt--ccccgagctcaccagtgaaaagagcctccagggctatatctactgctggtagt- 844
Db 35801 TTTTATTAAAGAACTCCCAAGAGAAAGGCGCCACCAAGACCATCTTCTGCTGCTGAGTT 35742
Qy 845 -ttgtcttcacatactgggtgactttacgtttctcatittcagggaggtgtacatgga 902
Db 35741 TCTTTGTGGTTATCTACTATGAGACTTAATGTCTCATCTCCACAAATGTTGTATGGG 35682
Qy 903 taaatgattctctgctagtggtgctcaggttatttggcccaatagctatgccgcaatta 962
Db 35681 TATTCAGCCCTGTCATCTACAGATGTCCACAAGTTTATGGTCAATGCCCTATGCCACTGTCA 35622
Qy 963 gtcctttgatgctaatttatgctgataaccaataattcaagactctgcaaatgtt 1017
Db 35621 GTCTATGCTGCTAATCAGATCTGATTAAGAATCATCATGATTCTGCCAAAGGT 35567

RESULT 14
AC073188
LOCUS      Homo sapiens BAC clone RP11-73B2 from 7, complete sequence.
DEFINITION AC073188
ACCESSION  AC073188
VERSION    AC073188.10 GI:15920117
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 65128)
            Sulston,J.E. and Waterston,R.
            Toward a complete human genome sequence
            Genome Res. 8 (11), 1097-1108 (1998)
99063792
REFERENCE  2 (bases 1 to 65128)
            Cordum,H. and Abbott,A.
            The sequence of Homo sapiens BAC clone RP11-73B2
            Unpublished (2001)
REFERENCE  3 (bases 1 to 65128)
            Waterston,R.H.
            Direct Submission
            Submitted (10-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            4 (bases 1 to 65128)
            Waterston,R.H.
            Direct Submission
            Submitted (04-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            5 (bases 1 to 65128)
            Waterston,R.
            Direct Submission
            Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Oct 4, 2001 this sequence version replaced gi:14647308.
            -----
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@wustl.wustl.edu
            -----
            Summary Statistics
            -----
            Center project name: H_NH0073B02

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-905H7, 2000 bp overlap; the clone sequenced to the right is RP5-1193P9, 2000 bp overlap. Actual start of this clone is at base position 36351 of RP11-73B2; actual end is at base position 72100 of RP5-1193P9.

Polymorphisms exist between AC073188 and AC006457. A single plasmid region exists between 68464-68496 and 68531-68559. A single stranded/chemistry region exists at base 68546. This clone contains an extra sequence fragment not represented in the finished neighbor, AC006457. This is submitted as H_NH0073B02.F1.

FEATURES

Location/Qualifiers	source
1..65128	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="7"	
/map="7"	
/clone="RP11-73B2"	
406..1310	
/clone_lib="RPCI-11"	
/rpt_family="L1"	
1317..1547	
/rpt_family="L1"	
3257..3634	
/rpt_family="L1"	
3635..3925	
/rpt_family="MaLR"	
3930..4120	
/rpt_family="MaLR"	
4136..4334	
/rpt_family="Alu"	
4341..4488	
/rpt_family="Alu"	
4555..4753	
/rpt_family="L1"	
5597..5693	
/rpt_family="MER1_type"	
5802..6205	

```
repeat_region /rpt_family="L1"
188..9281
/rpt_family="MER53"
10528..10758
/note="similar to EST A1028332 (NID:g3245641) ov90g03.x1"
repeat_region /rpt_family="MaLR"
10576..10799
/rpt_family="MaLR"
10813..11168
/rpt_family="MaLR"
11452..11647
/rpt_family="MER2_type"
11646..11696
/rpt_family="n" (TATG)n
11701..11833
/rpt_family="Alu"
12521..13095
/note="similar to EST A1806860 (NID:g5393426) wf36b07.x1"
repeat_region /rpt_family="AT_rich"
13058..13087
/rpt_family="AT_rich"
13146..13198
/rpt_family="AT_rich"
13365..13671
/rpt_family="Alu"
14087..14115
/rpt_family="n" (TTTTTA)n
14129..14883
/rpt_family="L1"
15223..15246
/rpt_family="AT_rich"
15290..15571
/note="similar to EST AW445087 (NID:g6986849)"
15290..15571
/note="similar to EST AW627961 (NID:g7374751) hl36d12.x1"
16305..16374
/rpt_family="Alu"
16753..17021
/rpt_family="Alu"
17038..17286
/rpt_family="Alu"
17700..18467
/rpt_family="L1"
17853..17872
/rpt_family="n" (CAAAA)n
18468..18778
/rpt_family="Alu"
18752..18796
/rpt_family="n" (A)n
18779..18938
/rpt_family="L1"
18838..18870
/rpt_family="AT_rich"
18944..19246
/rpt_family="Alu"
19220..19265
/rpt_family="n" (GAAA)n
19256..19691
/rpt_family="ERVL"
19748..19790
/rpt_family="n" (CAAAA)n
20557..20613
/note="similar to EST AW445087 (NID:g6986849)"
20557..20666
/note="similar to EST A1028332 (NID:g3245641) ov90g03.x1"
20594..20828
/note="similar to EST AA292895 (NID:g1941876) zt66c08.r1"
21125..21428
/rpt_family="Alu"
21664..21691
/rpt_family="AT_rich"
22442..22750
/note="similar to EST BF899282 (NID:g12290741)"
22640..23050
/note="similar to EST A1279713 (NID:g3917947) q153f03.x1"
```

```
misc_feature 32855..23086
/note="similar to EST AA292895 (NID:g1941876) zt66c08.r1"
repeat_region 23402..23451
/rpt_family="T-rich"
repeat_region 23430..23733
/rpt_family="Alu"
repeat_region 23850..24148
/rpt_family="MER2_type"
repeat_region 24586..25150
/rpt_family="L1"
repeat_region 25151..25275

Query Match 25.0%; Score 278; DB 9; Length 65128;
Best Local Similarity 62.4%; Pred No. 2,9e-61;
Matches 552; Conservative 0; Mismatches 320; Indels 12; Gaps 7;

QY 140 ttattaccatacaaatcttctttttatccacaagcagcttttgggaatttcagcaaacacc 199
Db 12154 TGTATGTTATTAAAAAGCTGCTTTGTTTACAAGTTGCTGTTGGAATCTCAGTCAATACT 12213

QY 200 atccttcttctttccacattcttccaccccttggtttcagtcacaggtcctaagccaattgac 259
Db 12214 TTTCGTGTTCCCTTTCAGCATCTTCACCTCTCTCTCTGTATCAGACATAAAACCCACTGAC 12273

QY 260 atgataattagtcacctgtctctccacatactgctgctcttccactcagcggaatttg 319
Db 12274 TTGATCATCTGTCCTTGCCCTCGTCCACATAGTAGTAACCTCTTCAATGACATGTTATG 12333

QY 320 gtgtccttagacttcttgggttcacagaatactcaggatgatctttaggtataaggtcatt 379
Db 12334 CAGTCTCCAAATCTGTTGCATCATTTGAA-CTTCAGAAATGACTTCAAGTGAAGGACATT 12392

QY 380 gctcttttaaaaagtgatgaggggctctccactgcaacccctgctcctctgagtg 439
Db 12393 TTCCACCTGAGCAGGGGTGTGTGGGGCTCTCTATCTGACCACCTGCCCTCTGAGTGTG 12452

QY 440 ctccagg---ccatcatcagcccgagcatctctctcttgccaaagctcaaacatctctt 496
Db 12453 TTCCAGGCCACCACTACATGAGCCCTGCTGATTTGGCAAGGTTTAAACATAAATCT 12512

QY 497 gcaagtcacatcttaggattcttcttcttctcctatgggttcacacatgttctattggt-gt 555
Db 12513 TCAAATATCATTTATCATGTTTCTTCTTCTTATGTTGTTCTCAACTTGTCTTTTCACTAGT 12572

QY 556 aatctctgtgtacactcggtctacccccagtg-aaacggggcgagctctcttctgttgc 614
Db 12573 AACAGGATCTTCTACACTGTAGTCTGCTTCCAGTGTGAGCAGACAAAATGACTGAATATA 12632

QY 615 atacagcactgttctcttcttggccatgagctacacccacagagactgtttttcaacta 674
Db 12633 GGTAAGTACTGCTCAGTGTTCCTCCATGAGCTTCATCTCCTCAGGCCACGGTTCTTATTCTG 12692

QY 675 atgacttgagggtatgcaccttttatagggttcattgggtctcctcagaggctacatgggtg 734
Db 12693 ACAACACTTAGGGATGCTTTCTCTGGTAGGAATACGTCTTCTTAATTCATACATAGTGTG 12752

QY 735 atattttatcacagaacaataagagctctctcagtccttcacgagccagcagcagctgtccc 794
Db 12753 ACTCTCTTATTCAGGC-CTCAGAGGTGATCCCAACCTTCACAGCAATGGTCTTCTCC 12811

QY 795 --gagcttcaccagtgaagagcctcccgagctcttactctggtggtgagttt---tgt 849
Db 12812 AGGACTTCCCCTGAGAAACGGGCCATCCAGACCACTACTGCTGCTGGTAAGTTTCTCTCT 12871

QY 850 cttcacactctgggtgacctttacgttctcttctcatttcagagggtgtgacatggataaata 909
Db 12872 GGTCTCTCTGTTGGTGGATTTCATCATCTCTCTCTCTCCTCAGCAATGTTGTGGGCATATGG 12931

QY 910 tctctgtagtggtgctccaggttattgtggccaataagctatgcgcgaattagctcttt 969
Db 12932 CCCAATCAACCA3CAAGTTTCAGAGTCTTTGTAGTCAATGCTATGCGGTGATCAGTCTCT 12991

QY 970 gatgctaatttatgctgataaccacaaatattcaagactcttgcaaa 1013
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 02:40:43 ; Search time 2520.32 Seconds
(without alignments)
5965.754 Million cell updates/sec

Title: US-09-728-309-2
Perfect score: 1114
Sequence: 1 attccagatcatagagatgt.....gagaggttaatccatcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estma:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_estl:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	199.6	17.9	725	12	AQ078448
2	186.6	16.8	690	12	AZ840584
3	186.2	16.7	777	12	AZ741291
4	159.6	14.3	808	12	BH039878
5	153.2	13.8	729	12	BH083291
6	142.6	12.8	545	12	AZ748309
7	141.8	12.7	617	12	AZ795382
8	136.4	12.2	635	12	BH062438
9	133.8	12.0	574	9	A1806860
10	130.4	11.7	610	12	AZ222872
11	130	11.7	467	12	AQ117475
12	129.4	11.6	615	12	AZ553624
13	129	11.6	563	12	AZ810472
14	128.4	11.5	740	12	BH267496
15	127	11.4	638	12	AZ553591
16	120.6	10.8	853	12	AZ265296
17	119.8	10.8	493	12	AZ759046

18	116.4	10.4	447	9	AA012849
19	116.4	10.4	768	12	BH060668
20	115.4	10.4	597	12	AZ422002
21	115	10.3	715	12	AG013852
22	113.2	10.2	698	12	BH039875
23	112	10.1	643	12	BH048022
24	111.8	10.0	397	9	AA021085
25	111.8	10.0	586	12	AZ093307
26	111.2	10.0	424	10	H86939
27	110.2	9.9	717	12	AG013832
28	109.6	9.8	682	12	AG130179
29	105	9.4	511	12	B67083
30	104	9.3	495	12	AQ223006
31	100	9.0	243	12	AZ737781
32	98.4	8.8	716	12	AG013834
33	98	8.8	696	10	BE968588
34	97.8	8.8	715	12	AZ994117
35	97.2	8.7	725	12	AG013831
36	94.6	8.5	382	9	AA442630
37	94.4	8.5	677	12	BH267500
38	92.6	8.3	580	9	AW973537
39	90.4	8.1	481	12	B92737
40	88	7.9	593	12	AZ754210
41	86.8	7.8	597	12	AZ896363
42	86.2	7.7	406	12	AZ767149
43	85.6	7.7	658	12	AQ666848
44	84.6	7.6	207	12	B59457
45	83	7.5	675	12	AG000745

ALIGNMENTS

RESULT 1	AQ078448	725 bp	DNA	linear	GSS 20-AUG-1998
LOCUS	CIT-HSP-2358C21.TFB	CIT-HSP	Homo sapiens	genomic clone	2358C21, DNA
DEFINITION	sequence.				
ACCESSION	AQ078448				
VERSION	AQ078448.1	GI:3439632			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 725)				
AUTHORS	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.				
TITLE	Use of a random human BAC End Sequence Database for Sequence-Ready Map Building				
JOURNAL	Unpublished (1998)				
COMMENT	Other_GSSs: CIT-HSP-2358C21.TR Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 338 0200 Fax: 301 338 0208 Email: mdadams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html. Seq primer: M13-21 Class: BAC clones.				
FEATURES	Location/Qualifiers				
source	1..725				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="2358C21"				
	/clone_lib="CIT-HSP"				
	/sex="Male"				
	/cell_type="Sperm"				

Db 390 ATATCAAGTGCATCTTTTCTGAGTGTCTCTATATATGTTAATTGGCAGTCAACTCTTG 449

Qy 565 ctgtacactcggcgtacccccagtgaaacggggcgagctctctctgtt-tgtcaatcacagcac 623

Db 450 GTATCCATCATTTGGCACCACCCCAATTTGACCATGAATGATTTATTTATGTTACTACGCC 509

Qy 624 tttctcttttggccatgagctacacccacagagagactgtttttcacacataagactttg 683

Db 510 TGTCTATTCTACCTTGAGTTACTCTATGCAAGCATATATTTCTACACTGCTAGCCATC 569

Qy 684 agggatgtcaacttttagggttcatggtctctcaagagagctacatgggtgattatttta 743

Db 570 AAGGAATTTCTTCTATTAGTCTCATGCTCTCTCTAATTTGGTACATGTTAGCCCTCTTA 629

Qy 744 tacagacaataagaggctattctcagtccttcacacagcagcagctgccccga 796

Db 630 AGCATGC-ACAGGAACAGACACAGCATCTTCAATGGAACCAACCTGTGCCCA 681

RESULT 3

AZ741291/c 777 bp DNA linear GSS 25-JAN-2001

LOCUS AZ741291

DEFINITION RPCI-24-90L5.TV RPCI-24 Mus musculus genomic clone RPCI-24-90L5, DNA sequence.

ACCESSION AZ741291

VERSION AZ741291.1 GI:12516500

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 777)

AUTHORS Zhao S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P., and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPCI-24

JOURNAL Unpublished (1999)

COMMENT Other GSSs: RPCI-24-90L5.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html

Plate: 90 row: L column: 5

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. .777

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-90L5"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 260 a 141 c 154 g 222 t

ORIGIN

Query Match 16.7%; Score 186.2; DB 12; Length 777;

Best Local Similarity 61.1%; Pred. No. 2.1e-34;

Matches 389; Conservative 0; Mismatches 238; Indels 10; Gaps 5;

Qy 385 tttaacaagaatgatgaggggctctccatctgcacccccgtcctcgtgagtgctccca 444

Db 774 TTTAAACAGGGCAATGAGAGGTCCTCTATCAGCATCACCTGCCTCTAGAGTGTGCA 715

Qy 445 ggcctatc--atcagccccagcatcttctctggcaagctcaaacatctcttctgcaag 501

Db 714 GGCTGTGACGATCAGTCCCAACTCCGCTTTGTTGGCAATTTAAACATAAAACATA 655

Qy 502 tcacattctaggattctctctctctctctctcgcacatgttctcatttggtgtaattct 561

Db 654 ATACATGATCTATTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 595

Qy 562 ctgct--gtacactgcggtacccccagtgaaacggggcgagctctctctgtttgtccata 619

Db 594 GATCTTCTATACTGCTGCTTATATCAACGCTGAGTGCAGACCAACGAGGTGAAGGTC 535

Qy 620 gcactgtctctcttttgcctatgagctacacccacagagagactgtttttcacactaatgac 679

Db 534 ATACTGCTCACTCTTCCCTATGAACAACATCATCAGGGGACTAATTTCAACAATCATAAC 475

Qy 680 ttgagggatgtcacctttataggggttcctcctcctcctcctcctcctcctcctcctcct 739

Db 474 CTTAAGAGATCTATTTTGTAGGAGTCATGCTGAGCACAAGTACATACATGCTGATTAT 415

Qy 740 ttatcacagacaataagaggtctctcagtccttcacacagcagcagcctgtcccccagtc 799

Db 414 CTTCTTTCAGAG-ATCAGAGGCAATGCAAGCATCTTCATAGCGACAGCCACTTGA-GAGCC 357

Qy 800 tcaccagtgaagaagagctccctcaggtctctctctctctctctctctctctctctctctct 856

Db 356 TCCCTCTGAGAAAAGAGCCACCACCATCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 297

Qy 857 tactgggtgagctttacgttctctcattttctcagggaggttgacatggataaattctctct 916

Db 296 TACTGGGTGGACTTTATATGTCATCTAGAACAGTCTTTTATGATGATGATGATGATGATG 237

Qy 917 ctagtgtggtccaggtattgtggccaaatagctatccgcaattagctcctttcttactacta 976

Db 236 ATGCTTACTGTTTCAGAAAGTTGTGTTGAATGCCATATCCCAATTTACTCTTTTAAACAA 177

Qy 977 attatgtctgataacaaatattcaagactctgcaaa 1013

Db 176 ATTAGTTCTGATATCGAATACTCATGATTTCTAAAAA 140

RESULT 4

BH039878

LOCUS BH039878

DEFINITION RPCI-24-273F22.TV RPCI-24 Mus musculus genomic clone RPCI-24-273F22, DNA sequence.

ACCESSION BH039878

VERSION BH039878.1 GI:14818265

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 808)

AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P., and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPCI-24

JOURNAL Unpublished (1999)

COMMENT Other GSSs: RPCI-24-273F22.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org


```
QY 258 acatgataattagtcacccgtgtctctcattccatccatcactgctgctcttccactcaggcaatat 317
Db 253 ACCTGCCCATGCTCTCTCTCCCTAAATCCACCTCATGATGCTTCT---GGTCGAGCAT 309
QY 318 tgggtccttagactcttgggttcacagaaactcactcagatgatctttaggtataaggtca 377
Db 310 TCATAGCCACACACATTTTATCTCTCGGAGGGCTGGGATGACATCATATGTAAATTC 369
QY 378 ttgtctttttaaacaggtgtaggggctctccatctgcacccctgcctcctcagtg 437
Db 370 TTGTGTACCTGTACAGAGTTTGTAGGGGTCTCTCCCTTTGCACCCAGCATGTTGAGTG 429
QY 438 tctcagaccatca---tcacccagcagcatctctcttgcaaaagctcaaacatcctt 494
Db 430 TCTTCAGCCCATCATCTTATGCTGATTTCTCCAGAGCTCTCTGTTATCAAGTTCAAGCATATAT 489
QY 495 ctgcaagtcaactcttaggattctctcttctctcattggtctcacaatgttctcatt-ggt 553
Db 490 CTCTTCATCATCTTATGTGCTATTTCTCTGAGTGTCTCTATATGTTAAATTAGCA 549
QY 554 gtaactctctgtgtacactcgggtacccccagtgaaac---ggggccagctctctctgt 609
Db 550 GTCAACTCTTGTGTATCAATTCATTTGCCACCCCTAAATTTGACCAAAATGACCTTCCTACTT 609
QY 610 ttgtcatcagcactgttctcttcttcttctcattggtctcacaatgttctcatttctc 668
Db 610 ATGTTACTCAGTCTGCTCTATTTCTTCTGAGTGTCTCTATATGTTAAATTCT 569
QY 669 acactaatgactttgagggatgtcacctttatagggttctcctcctcagagagctac 728
Db 670 AGCGTGTACGCATCAGGGGATCTTCTTATTAGTCTCATGTTCTCTCAACTGGGTAC 729
```

RESULT 6

```
AZ748309
LOCUS
DEFINITION
  AZ748309
  RPCI-24-63G18-TV RPCI-24 Mus musculus genomic clone RPCI-24-63G18,
  DNA sequence.
ACCESSION
  AZ748309
VERSION
  AZ748309.1 GI:12533043
KEYWORDS
  GSS.
SOURCE
  house mouse.
  Mus musculus
  Other_GSSs: RPCI-24-63G18.TJ
  Contact: Shaying Zhao
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: szhao@tigr.org
  Clones are derived from the mouse BAC library RPCI-24. For BAC
  library availability, please contact Pieter de Jong
  (pdejong@mail.cho.org). Clones may be purchased from BACPAC
  Resources (http://www.choi.org/bacpac/orderingframe.html). BAC end
  page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
  Plate: 63 row: G column: 18
  Seq primer: T7
  Class: BAC ends.
```

FEATURES

```
source
  Location/Qualifiers
    1..545
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="RPCI-24-63G18"
```

```
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/Note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
```

BASE COUNT 160 a 116 c 112 g 157 t
ORIGIN

Query Match 12.8%; Score 142.6; DB 12; Length 545;
Best Local Similarity 61.1%; Pred. No. 6.1e-24;
Matches 334; Conservative 0; Mismatches 204; Indels 9; Gaps 6;

QY 301 ctctactcaggcaataattggtgtccttagactcttcttgcacagaatactcagatga 360
Db 1 CCTCACTGGAGGAAATGTCTATGCTTCGACACATATTTGAGTCACCTGAATGTTGAGAATGA 60

QY 361 tcttaggtataaggtcattgtctcttttaacaaggtgtaggggctctcctcattcgcac 420
Db 61 CATCAAAATGTAAGGCAACTTTTATACAAACAGGGGTGATGAGAGGCTCTCTATACGACAT 120

QY 421 cccctgcctcctgagtgctcctcaggc---catcatcagcccccagcatctctccttggc 477
Db 121 CACTGCTCTCTGAGTGTGATCCAGGCTGTCAACATCATCTCGAGTACCTTTATGTTGCGCA 180

QY 478 aaagctcaa-acatcctctctgcaagtcacatctcttaggattcttcttctctcatgggtcc 536
Db 181 AAATTTAAACAATAAATAAGAAACACATGGTCAATGCTTCTTTTATATATTTGGTCT 240

QY 537 tcaacatgt--tcattggtgtaattctctgtgtacactgcgggctacccccagtg--aaac 593
Db 241 TTCAATTTTCTCTAGTAGTAGTCTCTGATCTTCTATACGTGGTGTTTTACCAATGTGAGT 300

QY 594 gggggcagctctctgtttgtctacagcactgtctcttcttgcacatgagctacaccccac 653
Db 301 GAGCCACAGCATGAGATGACTAAGTCTCTCAATCTTGCCAATGAACTACATCATC 360

QY 654 agggactgtttttcacacataatgactttgaggatgtcaccttttataggggttcaggttc 713
Db 361 AGGGGAATGGTTGTAACAGTACACACCGTCAGAGATGTGTTCTTGTGGAGTTATGCTG 420

QY 714 ctctcaagggtacataggtgatttttatcacagacaataagaggctatctcagtgct 773
Db 421 ATCAACAAGTGCATACATGTTGATGATTATCTTTGTCAGAC-ATCAGAGGCAATGCAAGCATCT 479

QY 774 tcacacagcagcctgtctcccgagtcctcaccagtgaaaagagcctccaggtctatcttac 833
Db 480 TCATAGCATCA-ACAACCTGAGAACCTCCCTTGAGAAAAGGCCACACAGACCATCTTGC 538

QY 834 tgctggt 840
Db 539 TGCTAGT 545

RESULT 7

```
AZ795382/c
LOCUS
DEFINITION
  AZ795382
  2M0049N06R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
  clone UUGC2M0049N06 R, DNA sequence.
ACCESSION
  AZ795382
VERSION
  AZ795382.1 GI:12942351
KEYWORDS
  GSS.
SOURCE
  house mouse.
  Mus musculus
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
    1 (bases 1 to 617)
    Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
    Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
    M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
```

and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
JOURNAL
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0049 row: N column: 06
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 617.

FEATURES
source

1. .617
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG2M0049N06"
/clone_lib="Mouse 10Kb plasmid UUCG1M library"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 214 a 101 c 130 g 172 t
ORIGIN

Query Match 12.7%; Score 141.8; DB 12; Length 617;
Best Local Similarity 57.2%; Pred. No. 9.9e-24;
Matches 336; Conservative 0; Mismatches 242; Indels 9; Gaps 4;

QY 81 ttttccacagatattcttgccttaatttcttcttaagatgatcaacttctggtt 140
DB 584 TTTTTCAGATATGAATAGGCTCCATATGATTTTAAAGATGATACACTTCCATT 525
QY 141 ttat-taccatacaaatcttctttatcacagccagcttttggaatttcagcaaacacc 199
DB 524 ATGTCTTCAATCAAGATATACCTTTATTTCCAAAGCTGGTCTAGGAGTCTACGCCAATATG 465
QY 200 atcttcttcttccacatcttccacttcttgcagtcacaggtcttaagtcattgac 259
DB 464 TTTCTACTGACTTTCTATATTTTCAATCCCAT---GTCACAGTCTTAAGCTGTGGAC 408
QY 260 atgataattagtcacctgtctctatccacatactgctgtcttcaactcaggcaaatatig 319
DB 407 CTGATCTCTGCTCAACTGACTTTTATTTACATTAATATGCTCTCACTGGATGGGATATC 348
QY 320 gtgtccttagacttcttgggttcacagaatactcaggatgatcttaggtataaggtcatt 379
DB 347 TGGCTTACAGACATGTTTGAGTTACTGACATTTGAGATGACTTCAATGTAAGGCACT 288

QY 380 gtctttttaaacaagtgatgaggggcctctccatctgacccccctgctctctgagtggtg 439
DB 287 TTTTACATAAACAGGGGTGATGAGAGCCCTCTCTATCTGTCATCACCTGCCCCCTGAGTGTA 228
QY 440 ctccaggc---catcatcagcccccagcatcttctctctggaagcctcaaacatctcttct 496
DB 227 TTCCAGGCTGTCTCATATAGTCCCACTCTCTGTCAGTCTTGGCAAAATTTAAACATAAAGTG 168
QY 497 gcaagtcacatctaggattcttcttctctctctggtctcaacatgttctcatgtgtgta 556
DB 167 AAAAATAATACATATATGCTTTCTTCTATGTTTGGTCTTTCAATTTGCTTTCACAGT 108
QY 557 atcttctgcttacct--gcggctaccgccagtgaaagggggccagcttctctgtttgtc 614
DB 107 AACCAGTCTTCTCTGTTGGTCTTATGCCAATCTGAGTGAGACCCAGCAGATGAAGGTC 48
QY 615 atacagcactgttcttcttctgcccagctacacacccacaggagact 661
DB 47 ACTAATTACTGCTCACACTTTCCCATGAACAACATTTATAAGGGGATT 1

RESULT 8

BH062438
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
BH062438
RPCI-24-327B18.TJ RPCI-24 Mus musculus genomic clone RPCI-24-327B18
DNA sequence.
BH062438
BH062438.1 GI:14875948
GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 635)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Aknret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 327 row: B column: 18
Seq primer: SP6
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1. .635
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-327B18"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site:1; BamHI; Site:2; BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
174 a 130 c 125 g 206 t

Query Match 12.2%; Score 136.4; DB 12; Length 635;

Best Local Similarity 63.2%; Pred. No. 2e-22;
Matches 259; Conservative 0; Mismatches 146; Indels 5; Gaps 3;

Qy 607 tgtttgcatcacagcactgttccttggccatgagctacaccacagaggagactgttt 666
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 TGAAGGTGATTAAATACTGCTCACTCTTCCCATGNACTACATCATCAGGGACTGATTT 62

Qy 667 tcacctaataagtacttgaggagatgcacctttatagggtttcattggtctctcaagagct 726
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 TAACCATGACAACCTTAAGAGATGTGTTCTTGTAGGAGTCATCGTGAACGACAT 122

Qy 727 acatgggattatttttatcacagacaataagaggctatctcaagtgccttcacacagccaagc 786
||||| |||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 ACATGGGTGATTATCTGTGTCAGC-ATCAGAGGCAATGCTTTTTTCTTCATAGCATCAAC 181

Qy 787 ctgtcccccgagctccaccagtgaagaagagcctcccaggctatcttaactgctgggtgagt-- 844
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 C-ACCTGAGAGCCTCCCCTGAGAAAAGGCCACTCATACCATCTTGCTGCTGGTGGTTT 240

Qy 845 -tttgtttcacatactcgggtggactttaocttcbtcaatttcaggaggagtgtgacatgat 903
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 CTMTTGTGGTCATGTACTGTTGAGTCTTCATCATCTCATCCACCTCAGTCCCTGTATGGAT 300

Qy 904 aaatgattctctgtagdgtggctccagggttatgtggccaatagctatgccgaattag 963
||||| |||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GTATGACCCAGTCATCTGACTGTTTCAGAAGTTGTGATGAATGCCTATCTTACAAATTAC 360

Qy 964 tcctttgatgctaattatgctgtataccaaattattcaagactctgcaaa 1013
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 TCCATTGGTCAAAATCAGTCTTGATAGCGGAATATCATGATGCTAAAAA 410

RESULT 9
AI806860/c
LOCUS
DEFINITION
IMAGE:2357653 3' similar to TR:Q62855 Q62855 PHEROMONE RECEPTOR VN6. ; mRNA sequence.
AI806860
VERSION AI806860.1 GI:5393426
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 574)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs@email.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 659 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 447.
Location/Qualifiers
FEATURES
source 1..574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of

RESULT 12	AZ553624	615 bp	DNA	linear	GSS 20-NOV-2000
LOCUS	RPCI-23-209L1-TV	RPCI-23	Mus musculus	genomic clone	RPCI-23-209L1, DNA sequence.
DEFINITION	AZ553624				
ACCESSION	AZ553624.1	GI:11233124			
VERSION					
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 615)				
AUTHORS	Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroi, M., de Jong, P. and Fraser, C.M.				
TITLE	Mouse BAC End Sequences from Library RPCI-23				
JOURNAL	Unpublished (1999)				
COMMENT	Other_GSSs: RPCI-23-209L1.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html Plate: 209 row: L column: 1 Seq primer: T7 Class: BAC ends.				
FEATURES	source				
	1..615				
	/organism="Mus musculus"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="RPCI-23-209L1"				
	/clone_lib="RPCI-23"				
	/sex="Female"				
	/lab_host="DH10B"				
	/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site:1: EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."				
BASE COUNT	178 a 128 c 107 g 202 t				
ORIGIN					
Query Match	11.6%;	Score 129.4;	DB 12;	Length 615;	
Best Local Similarity	58.3%;	Pred. No. 9.8e-21;			
Matches 284;	Conservative 0;	Mismatches 196;	Indels 7;	Gaps 3;	
QY	99	tttgcttaatttcttctaagatacatacaacttctctggtttttattaccatacaaatct	158		
Db	80	TATGACTCAACAATGCTCTTAAGATAAATTCACCTCTCACTTATGTCTCATTTGGAGAAT	139		
QY	159	tct-tttatccacagccagcttggaatttcagcaaacaccatctcttctttccac	217		
Db	140	GCTCTTATATCCAACTGGGTTAGAGAGCTTACCTAGCAATATGTGCTTCTTGTTTCTAT	199		
QY	218	atttcacctttgttttcagtcacaggttcagatccatgacatgataatagcacctg	277		
Db	200	ATTTTCA---TAATCTAGGGCAGACACCTAAGCCCATGACCTAATCTCTCTCAACAG	256		
QY	278	ttctcatcacatactgctgctcttcaactcaggcaaatattgcttcttagacttcttt	337		
Db	257	ACTTTCAATACATAATGCTGTTCTTCACTGCAGGAGATATTTTGATACAGATATTTT	316		
QY	338	ggttcagaataactcaggatgatcttagtataaagtcattgtctttttaacaagggtg	397		
Db	317	GAGTCAATGAATATTTGAGAATGACTTCAAAATGCAAAACAACATTTTACATATGCAGGTA	376		
QY	398	atgaggggcctctccatctgcacccctgctcctcctgaagtgtgctccagc---catac	454		
Db	377	ATGAGAGGGCTCTCTATCTGCACCACCTGCCTCTGAGTGTGTCCAGGTGTCCACCATC	436		
QY	455	agcccgagcatcttctctctggcgaagctcaaacatctcttgcgaagtcacatcttagga	514		
Db	437	AGTCCCAATACCTTCGCTGTGGCAAAATTTAAACATAAAATAAATAACACTATCAA	496		
QY	515	ttctctcttctctcaggtctcctcaacatgttcatgttggtgtaattcttctgctgtaactg	574		
Db	497	TGCTTTCTCTATATTTGGGTCTTTAAATTTGCTTCACTAGTAACCTCATCTCTCTATG	556		
QY	575	cggctac	581		
Db	557	TTGGTGC	563		
RESULT 13	AZ810472	563 bp	DNA	linear	GSS 20-FEB-2001
LOCUS	2M0D76E06F	Mouse 10kb	plasmid	UUGC1M library	Mus musculus genomic clone UUGC2M0076E06 F, DNA sequence.
DEFINITION	AZ810472				
ACCESSION	AZ810472				
VERSION	AZ810472.1	GI:12977756			
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 563)				
AUTHORS	Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D. Weiss, R.				
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0076 row: E column: 06 Seq primer: CGTTGTAAACGACGCGCCAGT Class: plasmid ends High quality sequence stop: 563.				
FEATURES	Location/Qualifiers				
source	1..563				
	/organism="Mus musculus"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="UUGC2M0076E06"				
	/clone_lib="Mouse 10kb plasmid				
	/sex="Male"				
	/lab_host="E. Coli strain XL10-Gold, Ti-resistant, P-"				
	/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA				

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g114732114[gblAF129072.1], a copy-number inducible derivative of plasmid R1). The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT	ORIGIN
164 a	116 c 105 g 178 t

Query Match 11.6%; Score 129; DB 12; Length 563;
Best Local Similarity 61.6%; Pred. No. 1.2e-20;
Matches 257; Conservative 0; Mismatches 155; Indels 5; Gaps 3;

Qy 613 tcatacagcactgttcctttttgcccatgagctacacccacaggaagactatttttcacac 672

[illegible]

Db 9 TCACTAAATCCTGCTTACTCTTCGCGCATGAACTCCATCAGGTATTCATTTTACAG 68

[illegible]

QY 073 caacyacccccyagggatgccccccatccatagggcccccaagggccccacaggg / 32

Db 69 TGACAACCTCCAGAGATGTATTTCTTGTAGGAATGATGCTAACCACAAGTACATACATGG 128

Qy 733 tgattattttacagacaataagaggctatctcagtgcccttcacacagccagcctgtcc 792

nb 129 TCATTCATCTCTACAC - ATCCACCCACCTCCACACCTTCATACCATCAC 186

[illegible]

Qy 793 ccgagtccaccagtgaagaagagcctccaggctatcttactgctggtgagt---tttgt 849

D5 187 GAGAGCATCCCCCTGAGAAAAGAGGCCACCCAGACTATCTTGCTGCTAGTGATTTTCTTTGT 246

850 cttcacatactggaatggactttacggtctctcattttcaggaggtatgcacatgcataaatga 909

RESULT	14
BH267496/c	
LOCUS	BH267496
DEFINITION	CH230-82L13.TVB Segment 1 Rattus norvegicus genomic clone CH230-82L13, DNA sequence.
	740 bp
	DNA
	linear
	GSS 30-NOV-2001

BH267496
 BH267496.1 GI:17179806
 GSS.
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 740)
 Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn
 A., Gerecseorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
 Jong, P. and Fraser, C.M.
 Rat BAC End Sequences from Library CHORI-230 EcoRI segment
 Unpublished (1999)
 Other_GSSs: CH230-82L13.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@email.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/orering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 82 row: L column: 13
Seq primer: T7
class: BAC_end

FEATURES	source
Class: BAC ends:	
Location/Qualifiers	
1. .740	
/organism="Rattus norvegicus"	
/strain="BN/SSNHsd/MCW"	
/db_xref="taxon:10116"	
/clone="CH230-82L13"	
/clone_lib="CHORI-230 Segment 1"	
/sex="Female"	
/cell_type="Brain"	
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"	
250 a 136 c 136 g 218 t	
BASE COUNT	
ORIGIN	

Query Match 11.5%; Score 128.4; DB 12; Length 740;
Best Local Similarity 64.9%; Pred. No. 1.8e-20;
Matches 253; Conservative 0; Mismatches 131; Totals 6; Cans 4.

627 tcccttttgcacatgagctacacccacaggaactgtttttcacactaatgacttttgaag 686

[illegible]

Db 740 TCACTCTTCCCATGAACATCATCATCAGGGACTGATTTAAACAATGACAACCTCCAGA 681

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

QY 68/ gatgtcacctttataggggttcattggtccctcaagaggctacatgggtgattattttatc /46

Db 680 GATGTTTCTA-GTAGGAGTCATGCTGACAACAAGCATGTACATGGTGATTATCTTGTGC 622

Qy 747 agacaataagaggctatctcagtcgcttcacacagccagcctgtccccgagtcaccag 806

631 ACAC-ATCCAGCCACCTCAGCGCATCTTTCATTACCACTCTCATCT-TGACACACTCCCCCTC

db

[illegible]

Qy 807 tgaaagagcctccaggctatctactgctggtgag---ttttgtcttcacatactggg 863

D5 563 AGAAAAGGGCCACCCAGACCATCTTGCTGTGGTGGGTTTTTTTGTGGTCATGTACTGGG 504

Ov 864 tggactttacgattctcatcttttcaggaggtatgacatggataaattgattctctcctagtgt 923

RESULT_15	AZ553591	638 bp	DNA	linear	GSS 20-NOV-2000
LOCUS	AZ553591				
DEFINITION	RPC1-23-209J1.TV RPC1-23 Mus musculus genomic clone RPC1-23-209J1, DNA sequence.				
ACCESSION	AZ553591				
VERSION	AZ553591.1	GI:112333013			
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				

Mus musculus
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

Mus musculus
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 638)
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other_GSSs: RPT-23-209JT:1J
Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9/12 Medical Center
Tel.: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPC1-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 209 row: J column: 1

Seq primer: T7

Class: BAC ends.

1: 638
LOCATION/QUALITY

/organism="Mus musculus"

```
/strain="C57BL/6J"
```

```
00T:HOXB7-1aTx_gp/  
/clone="RPCT-23-209  
/clone="RPCT-23-209
```

```
/clone_lib="RPCI-23"
```

```
/sex="Female"
```

```

/lab_host="DHIOB"
/note="Organ: Kidney"

```

7000-
Organ: kidney
EcoRI; Site_2: EcoR

brain genomic DNA

selected DNA was cloned into the pBac3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies).

184 a	135 c	105 g	214 t
-------	-------	-------	-------

Query:

Best Local Similarity 59.38; Pred. No. 3.8e-20;
Matches 272; Conservative 0; Mismatched 190; Indel

Matches 2/2; conservative 0; mismatches 100; indels

Qy 99 ttgctttaattttccttcttaagatgatcaaaccttcctgggtttattaccatacaaatct 158

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

50 39 TAAGACAGACAAAGCGCTCTAAAGATAAATTCACCTCCTCAGCTTATGTTCTTCATGGAGAA 98

Qy 159 tct-ttatccaagccagctttggaatttcagcaaacaccatccttcttcttccac 217

Db 99 GCCTCTTATATCCCAAGCTGGGTAGGAGTCCTAGCTAATAATATGTCCTTCCTGTTTCTAT 158

Qy 218 atcttcacctttgttttcagttcacagqctctaagttccattqacatgataattagtcacctg 277

[illegible]

Db 159 ATTTCA---TAATCCTAGGGCACAGACCTAAGCCCATGGACCTAATCTCTGTCAACAG 215

Ov 278 tctctcatccacatactactactctcttcaactcaggaataattggtatgctccttagactcttt 337

[illegible]

Db 216 ACTTTCATTCACATAATGCTGTCTTCTCACTGCAGGAGATATTTTGCATACAGATATATT 275

Qy 338 qdtrcacagaataactcacgaatgatacttagatataaaggtcatctctcttttaaacaaagctc 397

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 03:56:34 ; Search time 107.09 Seconds
(without alignments)
2555.194 Million cell updates/sec

Title: US-09-728-309-2
Perfect score: 1114
Sequence: 1 attccagatcatagagatgt.....gagaggttaattccattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCRUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62.4	5.6	7218	1	US-08-232-463-14
2	56.2	5.0	8353	3	US-08-611-587-1
3	48.4	4.3	246240	2	US-08-724-394A-20
4	48.4	4.3	246240	2	US-08-724-394A-21
5	48.4	4.3	246240	2	US-08-724-394A-22
6	47.4	4.2	638	4	US-09-328-111-275
7	44.4	4.0	497	4	US-09-328-111-109
8	43.3	3.9	152331	3	US-09-128-155-16
9	40.6	3.6	2270	2	US-08-852-807-16
10	40.6	3.6	13674	2	US-08-852-807-1
11	40.6	3.6	17410	1	US-07-841-646-3
12	40.6	3.6	17410	1	US-08-147-023-3
13	40.6	3.6	17410	1	US-08-447-570-3
14	40.6	3.6	17410	2	US-08-449-700-3
15	40.6	3.6	17410	2	US-08-449-699A-3
16	40.6	3.6	17415	3	US-08-486-343A-1
17	40.6	3.6	17415	5	PCT-US95-07349-1
18	39.3	3.5	504	4	US-09-328-111-120
19	36.2	3.2	80246	4	US-09-078-294-4
20	36.2	3.2	80595	4	US-09-078-294-3
21	35.4	3.2	289	4	US-09-007-005-17
22	35.4	3.2	289	4	US-09-244-796-17
23	35.2	3.2	2358	3	US-09-022-983-1
24	35.2	3.2	2360	4	US-09-490-692-10
25	35.2	3.2	40352	3	US-08-846-111B-15
26	34.2	3.1	4002	3	US-09-356-952-9
27	33.6	3.0	437	4	US-08-576-202-29

28	33.6	3.0	437	5	PCT-US95-16766-29	Sequence 29, Appl
29	33.6	3.0	3624	1	US-07-951-715A-6	Sequence 6, Appl
30	33.6	3.0	3624	2	US-08-459-448A-6	Sequence 6, Appl
31	33.6	3.0	3624	3	US-08-459-595A-6	Sequence 6, Appl
32	33.6	3.0	3624	3	US-08-459-504B-6	Sequence 6, Appl
33	33.6	3.0	3624	3	US-08-459-44A-6	Sequence 6, Appl
34	33.6	3.0	3624	3	US-09-053-549-7	Sequence 7, Appl
35	33.6	3.0	3624	4	US-08-547-422-6	Sequence 6, Appl
36	33.2	3.0	949	3	US-08-714-918-12	Sequence 12, Appl
37	33.2	3.0	949	4	US-09-265-315-12	Sequence 12, Appl
38	33.2	3.0	949	4	US-09-265-315-12	Sequence 12, Appl
39	33.2	3.0	949	4	US-09-266-417-12	Sequence 12, Appl
40	33.2	3.0	2529	2	US-08-867-129-1	Sequence 1, Appl
41	33.3	3.0	2518	4	US-09-433-699-3	Sequence 3, Appl
42	33.3	3.0	29598	4	US-09-341-587-6	Sequence 6, Appl
43	32.6	2.5	1307	1	US-08-229-418-1	Sequence 1, Appl
44	32.6	2.5	1307	2	US-08-932-761A-1	Sequence 1, Appl
45	32.6	2.9	1307	4	US-09-307-912-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgpt-F1s
: US-08-232-463-14


```
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONFIG"
US-08-724-394A-20

Query Match 4.3%; Score 48.4; DB 2; Length 246240;
Best Local Similarity 83.3%; Pred. NO. 0.0016;
Matches 55; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1011 aaatgttatggtttaaatattgtctctccaaagctcatgttgaaatttaacgccaat 1070
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 142977 AAGTGATATGTTGGATATTTGTCCTCTAAATCTCATGTGTAATCTCAAT 142918

Qy 1071 gtggca 1076
||| |
Db 142917 GTTGA 142912

RESULT 5
US-08-724-394A-22/c
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONFIG"
US-08-724-394A-22

Query Match 4.3%; Score 48.4; DB 2; Length 246240;
Best Local Similarity 83.3%; Pred. NO. 0.0016;
Matches 55; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1011 aaatgttatggtttaaatattgtctctccaaagctcatgttgaaatttaacgccaat 1070
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 142977 AAGTGATATGTTGGATATTTGTCCTCTAAATCTCATGTGTAATCTCAAT 142918

Qy 1071 gtggca 1076
||| |
Db 142917 GTTGA 142912

RESULT 4
US-08-724-394A-21/c
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONFIG"
US-08-724-394A-21

Query Match 4.3%; Score 48.4; DB 2; Length 246240;
Best Local Similarity 83.3%; Pred. NO. 0.0016;
```

Db 142917 GTTGA 142912

RESULT 6

US-09-328-111-275/c

; Sequence 275, Application US/09328111

; Patent No. 6262333

; GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.

; APPLICANT: Steinmann, Kathleen E.

; APPLICANT: Astle, Jon H.

; APPLICANT: Burgess, Christopher C.

; APPLICANT: Bushnell, Steven E.

; APPLICANT: Carroll III, Eddie

; APPLICANT: Catino, Theodore J.

; APPLICANT: Derti, Adnan

; APPLICANT: Ford, Donna M.

; APPLICANT: Lewis, Marcia E.

; APPLICANT: Monahan, John E.

; APPLICANT: Schlegel, Robert

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; FILE REFERENCE: CCD-257 (US)

; CURRENT APPLICATION NUMBER: US/09/328,111

; CURRENT FILING DATE: 1999-06-08

; EARLIER APPLICATION NUMBER: US 60/088,801

; EARLIER FILING DATE: 1998-06-10

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 275

; LENGTH: 638

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(638)

; OTHER INFORMATION: n = A,T,C or G

US-09-328-111-275

Query Match

Best Local Similarity 4.2%; Score 47; DB 4; Length 638;

Mismatches 62; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 990 accaaattattcaagactgcgaattgttgaatttgaatttgcctcctccaaagctca 1049

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 335 ACAAGTTCCTCTCCACCATCTGATGTTGGATGCTCTGCCCTCCAAATCTCA 276

Qy 1050 tgttgaatttaatcgccaatgtggca 1076

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 275 TGTGAAATGTAATTCACAGTGTGGA 249

RESULT 7

US-09-328-111-109/c

; Sequence 109, Application US/09328111

; Patent No. 6262333

; GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.

; APPLICANT: Steinmann, Kathleen E.

; APPLICANT: Astle, Jon H.

; APPLICANT: Burgess, Christopher C.

; APPLICANT: Bushnell, Steven E.

; APPLICANT: Carroll III, Eddie

; APPLICANT: Catino, Theodore J.

; APPLICANT: Derti, Adnan

; APPLICANT: Ford, Donna M.

; APPLICANT: Lewis, Marcia E.

; APPLICANT: Monahan, John E.

; APPLICANT: Schlegel, Robert

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; FILE REFERENCE: CCD-257 (US)

US-09-328-111-109

; CURRENT APPLICATION NUMBER: US/09/328,111

; EARLIER FILING DATE: 1999-06-08

; EARLIER APPLICATION NUMBER: US 60/088,801

; EARLIER FILING DATE: 1998-06-10

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 109

; LENGTH: 497

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(497)

; OTHER INFORMATION: n = A,T,C or G

US-09-328-111-109

Query Match 4.0%; Score 44.4; DB 4; Length 497;

Best Local Similarity 77.1%; Pred. No. 0.00095;

Mismatches 54; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1008 tgcgaatgttatggttaaatattgtctcctcctcaagctcatgttgaaatttaacgcc 1067

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 319 TCCTACTGATATGTTGGATGTTTGTCCGCTCCAAATCTCATGATGAATGTAATCCCC 260

Qy 1068 aatgtggcag 1077

||||| |||

Db 259 AGTGTGGAG 250

RESULT 8

US-09-128-155-16

; Sequence 16, Application US/09128155

; Patent No. 6117654

; GENERAL INFORMATION:

; APPLICANT: Pan, Yang

; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY

; FILE REFERENCE: 09404/052001

; CURRENT APPLICATION NUMBER: US/09/128,155

; CURRENT FILING DATE: 1998-08-03

; EARLIER APPLICATION NUMBER: US 60/091,650

; EARLIER FILING DATE: 1998-07-02

; EARLIER APPLICATION NUMBER: US 60/054,646

; EARLIER FILING DATE: 1997-08-04

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 16

; LENGTH: 152331

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(152331)

; OTHER INFORMATION: n = A,T,C or G

US-09-128-155-16

Query Match

Best Local Similarity 3.9%; Score 43; DB 3; Length 152331;

Mismatches 52; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1012 aatgttatggttttaaatattgtctcctcctcaagctcatgttgaaatttaacgccaatg 1071

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 73474 aataatggtttggatgtttgccccctccaaatctcacattgaaatgtaatcccccaatg 73533

Qy 1072 tggcagt 1078

|||||

Db 73534 ttgaagt 73540

RESULT 9

US-08-852-807-16/c

```
; Sequence 16, Application US/08852807
; Patent No. 5861298
; GENERAL INFORMATION:
; APPLICANT: Debouck, Christine
; APPLICANT: Drake, Fred
; APPLICANT: Gowen, Maxine
; APPLICANT: Rood, Julie
; APPLICANT: Hastings, Gregg
; APPLICANT: Adams, Mark
; APPLICANT: Fraser, Claire
; APPLICANT: Lee, No. 5861298man
; APPLICANT: Kirkness, Ewen
; APPLICANT: Blake, Judith
; APPLICANT: Fitzgerald, Lisa
; TITLE OF INVENTION: CATHEPSIN K GENE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,807
; FILING DATE: 07-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/019,942
; FILING DATE: 14-JUNE-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/020,273
; FILING DATE: 17-JUNE-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/026,273
; FILING DATE: 26-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50006-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2270 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-852-807-16

Query Match 3.6%; Score 40.6; DB 2; Length 2270;
Best Local Similarity 77.8%; Pred. No. 0.028;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1015 gttatggtttaatttgcctcccaagctcatgttgaaatttaatcgcgaatg 1074
| ||||| || |||| |||| | |||| |||| | || |||||
Db 576 GATATGGTTGGATGTGTGTCCTCCAGTCTCATGCTGAAATGTGATCCCAATGG 517
| ||||| || |||| |||| | |||| |||| | || |||||
Qy 1075 caq 1077
||
```

```
Db 516 GAG 514

RESULT 10
US-08-852-807-1/c
; Sequence 1, Application US/08852807
; Patent No. 5861298
; GENERAL INFORMATION:
; APPLICANT: Debouck, Christine
; APPLICANT: Drake, Fred
; APPLICANT: Gowen, Maxine
; APPLICANT: Rood, Julie
; APPLICANT: Hastings, Gregg
; APPLICANT: Adams, Mark
; APPLICANT: Fraser, Claire
; APPLICANT: Lee, No. 5861298man
; APPLICANT: Kirkness, Ewen
; APPLICANT: Blake, Judith
; APPLICANT: Fitzgerald, Lisa
; TITLE OF INVENTION: CATHEPSIN K GENE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,807
; FILING DATE: 07-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/019,942
; FILING DATE: 14-JUNE-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/020,273
; FILING DATE: 17-JUNE-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/026,273
; FILING DATE: 26-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50006-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-852-807-1

Query Match 3.6%; Score 40.6; DB 2; Length 13674;
Best Local Similarity 77.8%; Pred. No. 0.072;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1015 gttatggtttaatttgcctcccaagctcatgttgaaatttaatcgcgaatg 1074
```

Db 10826 GATATGTTGGATGTGTGTCCTCCCGAGTCTGATCTGAAATGTGATCCCAATGTGG 10767

Qy 1075 cag 1077

Db 10766 GAG 10764

RESULT 11

US-07-841-646-3
: Sequence 3, Application US/07841646
: Patent No. 5266683
: GENERAL INFORMATION:
: APPLICANT: OPPERMANN, HERMANN
: APPLICANT: OZKAYNAK, ENGIN
: APPLICANT: KUBERASAMPATH, THANGAVEL
: APPLICANT: RUEGER, DAVID C.
: APPLICANT: PANG, ROY H.L.
: TITLE OF INVENTION: OSTEOGENIC DEVICES
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TESTA, HURWITZ & THIBEAULT
: STREET: 53 STATE STREET
: CITY: BOSTON
: STATE: MASSACHUSETTS
: COUNTRY: U.S.A.
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/841,646
: FILING DATE: 19920221
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 810,560
: FILING DATE: 20-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 827,052
: FILING DATE: 28-JAN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 660,162
: FILING DATE: 22-FEB-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 621,988
: FILING DATE: 04-DEC-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 621,849
: FILING DATE: 04-DEC-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 616,374
: FILING DATE: 21-NOV-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 600,024
: FILING DATE: 18-OCT-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 599,543
: FILING DATE: 18-OCT-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 579,865
: FILING DATE: 07-SEP-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 569,920
: FILING DATE: 20-AUG-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 483,913
: FILING DATE: 22-FEB-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 422,613
: FILING DATE: 17-OCT-1989
: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 315,342
: FILING DATE: 23-FEB-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 232,630
: FILING DATE: 15-AUG-1988
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 179,460
: FILING DATE: 08-APR-1988
: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER, EDMUND R.
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: CRP-001CP6
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/248-7000
: TELEFAX: 617/248-7100
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 17410 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: homo sapiens
: FEATURE:
: NAME/KEY: exon
: LOCATION: 3192..3730
: OTHER INFORMATION: /label= EXON-1
: OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 10413..10414
: OTHER INFORMATION: /label= GAP-1
: OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
: OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
: OTHER INFORMATION: THIS SEQUENCE."
: FEATURE:
: NAME/KEY: exon
: LOCATION: 10696..10891
: OTHER INFORMATION: /label= EXON-2
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 10960..10961
: OTHER INFORMATION: /label= GAP-2
: OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
: OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN
: OTHER INFORMATION: THIS SEQUENCE."
: FEATURE:
: NAME/KEY: exon
: LOCATION: 11059..11211
: OTHER INFORMATION: /label= EXON-3
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 11351..11352
: OTHER INFORMATION: /label= GAP-3
: OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
: OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
: OTHER INFORMATION: THIS SEQUENCE."
: FEATURE:
: NAME/KEY: exon
: LOCATION: 11420..11617
: OTHER INFORMATION: /label= EXON-4
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 11721..11722
: OTHER INFORMATION: /label= GAP-4
: OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
: OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
: OTHER INFORMATION: THIS SEQUENCE."
: FEATURE:
: NAME/KEY: exon
: LOCATION: 13354..13436
: OTHER INFORMATION: /label= EXON-5


```
;
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
;
; US-07-841-646-3
;
; Query Match 3.6%; Score 40.6; DB 1; Length 17410;
; Best Local Similarity 77.8%; Pred. No. 0.082;
; Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
;
Qy 1030 ttgtctctccaaagtcattgttgaaatttaacgcgaatgtgacgactaagaagtg 1089
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8812 TTTGTCCTACAAAACATCGTTGAAATTTAAATGTCAGTGTAACATTATTGAGAGTT 8871
|
Qy 1090 atg 1092
|||
Db 8872 ATG 8874
|
;
; RESULT 12
; US-08-147-023-3
; Sequence 3, Application US/08147023
; Patent No. 5468845
; GENERAL INFORMATION:
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02109
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,023
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 530
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
;
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 579,865
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; THIS SEQUENCE."
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10636..10891
; OTHER INFORMATION: /label= EXON-2
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10960..10961
; OTHER INFORMATION: /label= GAP-2
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; BE MISSING BETWEEN POSITION 10960 AND 10961 IN
; THIS SEQUENCE."
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11059..11211
; OTHER INFORMATION: /label= EXON-3
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11351..11352
; OTHER INFORMATION: /label= GAP-3
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
; THIS SEQUENCE."
;
```

```

; OTHER INFORMATION: THIS SEQUENCE.
; NAME/KEY: exon
; LOCATION: 11420..11617
; OTHER INFORMATION: /label= EXON-4
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11721..11722
; OTHER INFORMATION: /note= GAP-4
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 13354..13436
; OTHER INFORMATION: /label= EXON-5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
; US-08-147-023-3
;
; Query Match 3.6%; Score 40.6; DB 1; Length 17410;
; Best Local Similarity 77.8%; Pred. No. 0.082;
; Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
;
; QY 1030 ttgtctcctccaaagctggttgaatttaatcgccaatgtggcagtaactaagaagtg 1089
;      ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8812 TTTGTCCTCAAAAACATCATGTGAAATTAAATGTCACGTGAACATTTATGAGAGTT 8871
;
; QY 1090 atg 1092
;      |||
; Db 8872 ATG 8874
;
; RESULT 13
; US-08-447-570-3
; Sequence 3, Application US/08447570
; Patent No. 5714589
; GENERAL INFORMATION:
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,570
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
;
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 579,865
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10696..10891
; OTHER INFORMATION: /label= EXON-2
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: 10960..10961
; OTHER INFORMATION: /label= GAP-2
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10960 AND 10961 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11059..11211
; OTHER INFORMATION: /label= EXON-3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11351..11352
; OTHER INFORMATION: /label= GAP-3
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11420..11617
; OTHER INFORMATION: /label= EXON-4
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11721..11722
; OTHER INFORMATION: /label= GAP-4
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 13354..13436
; OTHER INFORMATION: /label= EXON-5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
; US-08-447-570-3

Query Match          3.6%; Score 40.6; DB 1; Length 17410;
Best Local Similarity 77.8%; Pred. No. 0.082;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1030 ttgtctctccaaagctcatgttgaaatttaatcgccaatgtggcagtgactaagaagtg 1089
      ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8812 TTGTCCCTACAAAACATCATGTGTGAATTTAATTCACAGTGTAAACATTATTGAGAGGTT 8871

QY 1090 atg 1092
      |||
Db 8872 ATG 8874

RESULT 14
US-08-449-700-3
; Sequence 3, Application US/08449700
; Patent No. 5863758
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTROGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
```

```
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,700
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 579,865
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
```

```

; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10696..10891
; OTHER INFORMATION: /label= EXON-2
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10960..10961
; OTHER INFORMATION: /label= GAP-2
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10960 AND 10961 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11059..11211
; OTHER INFORMATION: /label= EXON-3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11351..11352
; OTHER INFORMATION: /label= GAP-3
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11420..11617
; OTHER INFORMATION: /label= EXON-4
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11721..11722
; OTHER INFORMATION: /label= GAP-4
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 13354..13436
; OTHER INFORMATION: /label= EXON-5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
; US-08-449-700-3
; Query Match 3.6%; Score 40.6; DB 2; Length 17410;
; Best Local Similarity 77.8%; Pred. No. 0.082;
; Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1030 ttgtctcccaagctcatgttgaaatttaaccccaatgtggcagtaagaagtg 1089
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8812 TTGTCCCTACAAAACACTCATGTGTAATTTAATTCAGTGTAACATTATGAGAGGTT 8871
QY 1090 atg 1092
Db 8872 ATG 8874
; RESULT 15
US-08-449-699A-3
; Sequence 3, Application US/08449699A
; Patent No. 5958441
; GENERAL INFORMATION:
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: ANTIBODIES TO OSTEOGENIC PROTEINS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,699A
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/147,023
; FILING DATE: 01-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: STR-001CP6CN
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10696..10891
; OTHER INFORMATION: /label= EXON-2
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10960..10961
; OTHER INFORMATION: /label= GAP-2
; OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10960 AND 10961 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11059..11211
; OTHER INFORMATION: /label= EXON-3
```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11351..11352
OTHER INFORMATION: /label= GAP-3
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 11420..11617
OTHER INFORMATION: /label= EXON-4
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11721..11722
OTHER INFORMATION: /label= GAP-4
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 13354..13436
OTHER INFORMATION: /label= EXON-5
FEATURE:
NAME/KEY: exon
LOCATION: 15044..15160
OTHER INFORMATION: /label= EXON-6
FEATURE:
NAME/KEY: exon
LOCATION: 17245..17410
OTHER INFORMATION: /label= EXON-7
US-08-449-699A-3
```

```
Query Match 3.6%; Score 40.6; DB 2; Length 17410;
Best Local Similarity 77.8%; Pred. No. 0.082;
Matches 49; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1030 ttgtctctccaaagctcattgtgaatttaatcgccaatgtggcagtaagaagtg 1089
||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8812 TTGTCCCTACAAAACTCAATGTTGAATTATTTGCCAGTGTAAACATTATTGAGCGTT 8871

Qy 1090 atg 1092
|||
Db 8872 ATG 8874
```

Search completed: July 21, 2002, 04:01:51
Job time: 7846 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2002, 03:57:56 ; Search time 376.86 Seconds
(without alignments)
5075.207 Million cell updates/sec

Title: US-09-728-309-2
Perfect score: 1114
Sequence: 1 attccagatcatagatgt.....gagaggttaattccattcatg 1114

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	229.8	20.6	1331	18 AAT69546	Rat pheromone rece
2	228	20.5	1538	18 AAT69549	Rat pheromone rece
3	220.2	19.8	1386	18 AAT69545	Rat pheromone rece
4	215	19.3	584	22 ABA60235	Human foetal liver
5	215	19.3	584	22 ABA28536	Probe #7002 for ge
6	215	19.3	584	22 AAK08511	Human brain expres
7	215	19.3	584	22 AAK34394	Human bone marrow
8	215	19.3	584	22 AAI40117	Probe #8803 used t
9	209	18.8	1496	18 AAT69547	Rat pheromone rece

10	207.4	18.6	1053	18 AAT69548	Rat pheromone rece
11	206	18.5	1265	18 AAT69550	Rat pheromone rece
12	186.2	16.7	1313	24 AAS98047	Human DNA for pote
13	183.2	16.4	1015	24 AAD24968	Human G-protein co
14	164.8	14.8	628	22 AAH74904	Probe used to isol
15	156.8	14.1	632	22 AAH74901	Probe used to isol
16	152.4	13.7	461	22 AAK06912	Human brain expres
17	152.4	13.7	461	22 AAK32640	Human bone marrow
18	152.4	13.7	461	22 AAK38455	Probe #7141 used t
19	149.4	13.4	632	22 AAH74899	Probe used to isol
20	147.8	13.3	628	22 AAH74900	Probe used to isol
21	145	13.0	653	22 AAH74903	Human purified sec
22	143.8	12.9	640	22 AAS63212	Human purified sec
23	139.6	12.5	662	22 AAH74902	Probe used to isol
24	137	12.3	636	22 AAH74905	Probe used to isol
25	134.8	12.1	300	22 ABA72774	Human foetal liver
26	134.8	12.1	300	22 ABA38413	Probe #16879 for g
27	134.8	12.1	300	22 AAK21203	Human brain expres
28	134.8	12.1	300	22 AAK47360	Human bone marrow
29	134.8	12.1	300	22 AAT53196	Probe #21882 used
30	133.8	12.0	574	24 AAS98151	Human DNA for pote
31	130.2	11.7	512	24 ABA05725	Rat pheromone rece
32	128.6	11.5	512	24 ABA05712	Rat pheromone rece
33	128.6	11.5	512	24 ABA05731	Rat pheromone rece
34	127	11.4	512	24 ABA05704	Rat pheromone rece
35	127	11.4	512	24 ABA05705	Rat pheromone rece
36	127	11.4	512	24 ABA05710	Rat pheromone rece
37	125.8	11.3	512	24 ABA05721	Rat pheromone rece
38	125.2	11.2	530	18 AAT69551	Human pheromone re
39	122.4	11.0	512	24 ABA05703	Rat pheromone rece
40	122.4	11.0	512	24 ABA05708	Rat pheromone rece
41	120.8	10.8	512	24 ABA05701	Rat pheromone rece
42	120.8	10.8	512	24 ABA05702	Rat pheromone rece
43	120.8	10.8	512	24 ABA05711	Rat pheromone rece
44	120.8	10.8	512	24 ABA05714	Rat pheromone rece
45	120.8	10.8	512	24 ABA05715	Rat pheromone rece

ALIGNMENTS

RESULT	1
AAT69546	
ID	AAT69546 standard; DNA; 1331 BP.
XX	AAT69546;
AC	AAT69546;
XX	26-AUG-1997 (first entry)
DT	
XX	Rat pheromone receptor VN3 cDNA.
DE	
XX	Pheromone receptor; vomeronasal sensory neuron; social behaviour;
KW	maternal behaviour; reproductive behaviour; fertility;
KW	hormone secretion; ss.
XX	
OS	Rattus sp.
XX	
FH	Key
CDS	Location/Qualifiers
FT	207..1142
FT	/*tag= a
FT	1038..1040
FT	/*tag= b
FT	/*transl_except= pos:1038..1040:_aa:Gly
FT	/note= "GTA codes for valine"
FT	1086..1088
FT	/*tag= c
FT	/transl_except= pos:1086..1088:_aa:Glu
FT	/note= "GGA codes for glycine"
XX	
PN	WO9714790-A1.
XX	
PD	24-APR-1997.
XX	

Db 424 gtaagtacagtaataactgatcaactttcttccataaagctacattatcaggagccgtgc 483
Qy 565 ttccacataatgactttgagggatgacacctttatagggttcattgtctctctcaagagg 724
Db 484 ttctatgctgccactgcttcaagatgtctctttgttagcaatcatcagccctcaagtgc 543
Qy 725 ctacatggtgattattttatcacagaca 751
Db 544 atacatggttaattctcttgttcaggca 570

RESULT 5
ABA28536
ID ABA28536 standard; DNA; 584 BP.
XX
AC ABA28536;
DT 23-JAN-2002 (first entry)
XX
DE Probe #7002 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
XX AAK08511 standard; DNA; 584 BP.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
XX Claim 1; SEQ ID No 7002; 530pp; English.
XX
PS The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 584 BP; 127 A; 173 C; 107 G; 177 T; 0 other;

Query Match 19.38; Score 215; DB 22; Length 584;
Best Local Similarity 64.78; Pred. No. 2.7e-50;
Matches 367; Conservative 0; Mismatches 195; Indels 5; Gaps 3;
Qy 190 agcaaacaccatctcttctttccacatcttcacattgtttttcagtcacaggtctaa 249

Db 4 agcaaacacttctctctattcttccaaaattcttccactcttcaggatcacagccctaa 63
Qy 250 gtccattgacatgataattagtcacctgtctctccacacatactgtctcttcaactca 309
Db 64 gccacagaccatcaccctgacccctgtccacttgcacacttaggatgctccctattgt 123
Qy 310 ggcataattggtgctcttagactcttcttggttccacagaatactcaggatgatcttagta 369
Db 124 ggtcttcttgcacatccagaccctgttggagtcactgtattttcagaatgactccaagt 183
Qy 370 taaggctcatgtctttttaaacaagggtgagggcctctcccatctgcacccctgct 429
Db 184 taaggcattcttctgcacaggggtgagggagcctctcccatctgcacccactgct 243
Qy 430 cctgagtggtctccaggc---catcatcagccccagcatcttctccttggcacaagctcaa 486
Db 244 cctgagcatgtccaggtgtgcgcacatcagccctggcacctcctggtctgcaaggattaa 303
Qy 487 acatccttctgcaagtcacatctttaggattctcttcttctcattggtctcacaacatgt- 545
Db 304 acagaaattcaaggttacatcttctcactctcttctctctatgggttctcagttgtc 363
Qy 546 tcattggttaattctctgtctgacactgcggctacccccagtgaaacggggccagtcct 605
Db 364 tctcagtagtaacctgtctctcactgtggtcttcttcttaattgggaccacagcggttgt 423
Qy 606 -ctgtttgtcatcacgacactgtctctttttgtcccatgagctacacccacagggagactgtt 664
Db 424 gctaaagtacagtaataactgatcacctttcttccataagctacattatcaggagccgtgc 483
Qy 565 ttccacataatgactttgagggatgacacacctttatagggttcattgtctctcacaagagg 724
Db 484 ttctatgctgccactgtcttcaagatgtcttcttctttagcaatcatcagccctcaagtgc 543
Qy 725 ctacatggtgattattttatcacagaca 751
Db 544 atacatggttaattctcttgttcaggca 570

RESULT 6
AAK08511
ID AAK08511 standard; DNA; 584 BP.
XX
AC AAK08511;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 8502.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
XX WO200157275-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;

Db 366 tgtctttcagtagtacataattttctcactgtggcttcttccattgtgacccagacca 425
Qy 601 gtctctgtgtgtacacagcagctgttctcttcttggccatgagtgatcacaccacagggagac 660
Db 426 atctacttaaggtccgcaaatcactgctcagctgttctcccatgaaatccatcatgtggggag 485
Qy 661 tgttttcacactaatgacttttgaggatgtgcacctttataggggttcattggtctctca 720
Db 486 tgttttcc-----ttgtaggattacgtgctctcaa 516
Qy 721 gaggtcatggtgattatttttacacagacaataagagggtctatctcagtgctcttcacaca 780
Db 517 gtgcatacatgatgatctttttgtcca-aggcatcagaagtgtatcccgacatcttcacagt 575
Qy 781 gcaagcctgtcccc--gagttccaccagtgaaagagagcctcccagggtatcttactgctg 838
Db 576 accagcctgtccccaagatctcgcagagaaaagggttaaccagatcatctgcccactg 635
Qy 839 gtgagt---tttgtcttcacatactgggtgacttttacgtctctcatatttcaggaggtgtg 895
Db 636 gtgaatgcttgtgtcatgttcttggtggagaccttatctatctatctcttcttccctg 695
Qy 896 acatggtataaatgattctctgtcgtagtgtggtccagggttattgtggccaatagctatgcc 955
Db 696 ttatggagctataaccagtcactcctgagcatctagaacctgtgtgctgtctatgcc 755
Qy 956 gcaattagctcttgatgctaattatgtcgtatataccaaatatcaagactctgcaaatg 1015
Db 756 actcgtcttcatttgggtacaaatccgctctgtgataaagaagtagtcaatattctccaaaa 815
Qy 1016 ttatggttttaaat 1028
Db 816 atggaattaaagt 828

RESULT 13
AAD24968
ID AAD24968 standard; cDNA; 1015 BP.
AC AAD24968;
XX 12-MAR-2002 (first entry)
DE Human G-protein coupled receptor-13 (GCREC-13) cDNA.
KW Human; G-protein coupled receptor-13; GCREC-13; therapy; cancer; stroke;
KW cell proliferative disorder; neurological; epilepsy; Parkinson's disease;
KW Alzheimer's disease; inflammation; thyroiditis; haemolytic anaemia; AIDS;
KW Acquired Immune Deficiency Syndrome; dementia; nontropic; cholelithiasis;
KW multiple sclerosis; atherosclerosis; angina pectoris; gastroenteritis;
KW diabetes; ulcer; viral infection; immunosuppressive; ss.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 248..1015
XX /*tag= a
XX /product= "Human GCREC-13 protein"
XX
XX W0200198351-A2.
XX
XX 27-DEC-2001.
XX
XX 15-JUN-2001; 2001WO-US19275.
XX
XX 16-JUN-2000; 2000US-212483P.
XX 22-JUN-2000; 2000US-213954P.
XX 29-JUN-2000; 2000US-215209P.
XX 07-JUL-2000; 2000US-216595P.
XX 14-JUL-2000; 2000US-218936P.
XX 19-JUL-2000; 2000US-219154P.
XX 21-JUL-2000; 2000US-220141P.
XX

PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lal P, Baughn MR, Hafalia AJA, Nguyen DB, Gandhi AR, Kallick DA;
PI Griffin JA, Yue H, Khan FA, Patterson C, Lu DAM, Tribouley CM;
PI Lu Y, Walia NK, Graul R, Yao MG, Yang J, Ramkumar J, Au-Young J;
PI Elliott VS, Hernandez R, Walsh RT, Borowsky ML, Thornton M, He A;
XX WPI: 2002-075627/10.
DR P-PSDB; AAE15643.
XX
PT Isolated human G-protein coupled receptor polypeptides and the use of
PT these sequences in the diagnosis, treatment and prevention of diseases
PT and in the assessment of exogenous compounds on the expression of the
PT receptors -
XX
XX Claim 11; Page 139; 143pp; English.
XX
CC The invention relates to isolated human G-protein coupled receptor
CC (GCREC) polypeptides and their biologically active fragments. GCREC and
CC protein is useful in treating a disease or condition associated with an
CC increase or decrease in expression of functional GCREC. The GCREC's are
CC useful in the diagnosis, treatment and prevention of cell proliferative
CC disorders (cancer, leukaemia, melanoma); neurological disorders (stroke,
CC epilepsy, Parkinson's disease, dementia, Alzheimer's disease); autoimmune
CC inflammatory disorder (thyroiditis, haemolytic anaemia, AIDS, multiple
CC sclerosis); cardiovascular disorder (atherosclerosis, angina pectoris),
CC gastrointestinal disorder (ulcer, cholelithiasis, gastroenteritis),
CC metabolic disorders (diabetes); viral infections (herpes virus) and in
CC the assessment of the effects of exogenous compounds on the expression
CC of the nucleic acid and amino acid sequences. The present sequence is
CC human GCREC-13 cDNA.
XX
SQ Sequence 1015 BP; 261 A; 226 C; 182 G; 346 T; 0 other;

Query Match 16.4%; Score 183.2; DB 24; Length 1015;
Best Local Similarity 59.1%; Pred. No. 3e-41;
Matches 469; Conservative 0; Mismatches 298; Indels 27; Gaps 8;
Qy 130 acttcctggtttttattaccatacaaatcttttttatccacaagcagcttggaaatttc 189
Db 222 actttgcacttgtgtcccatgaaaaatgcttctcatcttaagctggtattggcatctc 281
Qy 190 agcaaacacacatcctcttctttccacatcttcacctttg-----ttttcagtcaca 242
Db 282 agcaaacacactttctctcttcttctgtcgtctctctcttcttcttcttcttgatctca 341
Qy 243 ggtctaagtccattgacatgataattagtcacctgtctctctccacatactgtctct 302
Db 342 ggcagagcgcaacttaccatcaccagctgtcagtggtgcctcctccacatggtgctcttc 401
Qy 303 tcactcagggaataattggtgtccttagacttcttgggttcacagaataactcagatgac 362
Db 402 tcaccatgg---tgttctgtctccacagctcttgaactcactgaatttccagaatgact 458
Qy 363 ttaggataaggtcattgtcttttttaacaaggtgaggggctctccatctgcacccc 422
Db 459 tcaaatatgaggcatctcttctacctgaggagggtgacggagccttccatttggtaacca 518
Qy 423 cctgcctcctgagtggtctccag--gccatcacagcccccagcatcttctccttgccaa 479
Db 519 cctgcctcctcctggtggtcaggtcgtcgaacatcagccccagcatttccctgggtgga 578
Qy 480 agctcaaacacatccttctgcaagtccacatcttaggattcttcttcttctcaggggtcccca 539
Db 579 ggtttaaatggaaatccacaatttttacccttccattgttctcagtgctc-----tctcag 633
Qy 540 acatgttcaattggtgtaattcttctgtacactgcgggtaccctccagtgaaacggggccc 599
Db 634 ttttccctgttagtagtgactgtatcttttacactgtggtcttcttccaatgtgacccag-- 691
Qy 600 agtcttctgttctacacagcactgttcttcttcttcttcttcttcttcttcttcttcttctt 659

Db 692 atcaattgcatgctagtaataactgttcttccatttcccaataaaactccataatcagagga 751
Qy 660 ctgtttttcacacaaatgaattgaggagatgtcaccttttatagggttcattgctctctca 719
Db 752 ctgtttttcactctgatttattcagagatgtttttttaaacaataatgctgttctca 811
Qy 720 agagctcacatggtgattattttatcacagacaataaagaggtctatctcagtgccctcacac 779
Db 812 agtgtcacatgactctcattcaggaactacagagag--atcctggtaacctcaag 868
Qy 780 agccagcctgcccaggtctccacagtgaaaagag-cctcccaggtctatcttactgctg 838
Db 869 cccagcctctacctaagagatctttgcagaggaagccatcagcacatcctgctgccc 928
Qy 839 gtgagttt---tgtctcacatcctgggtgactttacgttctcattttcagagaggtgtg 895
Db 929 gtgagtttctcgggagtgatgaagatggacttcatctcaactctcacaactgtg 988
Qy 896 acatggataatga 909
Db 989 ccatgggcatatga 1002

RESULT 14
AAH74904
ID AAH74904 standard; DNA: 628 BP.
AC AAH74904;
XX
DT 29-OCT-2001 (first entry)
XX
DE Probe used to isolate human VNO cDNA sequences.
KW Human; vomeronasal organ; VNO; pheromone receptor; ion channel;
KW pheromone; vomeropherin; probe; ss.
XX
OS Homo sapiens.
XX
PN WO200161046-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05178.
XX
PR 17-FEB-2000; 2000US-0183128.
XX
PA (PHER-) PHERIN PHARM INC.
XX
PI Herman RC, Berliner D;
XX
DR WPI; 2001-522609/57.
XX
PT Novel vomeronasal organ cDNA library useful for identification and
PT isolation of genes expressed in vomeronasal organs, specifically genes
PT for pheromone receptors and ion channels -
XX
PS Example 9; Page 54-55; 59pp; English.
XX
CC The present sequence represents a probe, used to isolate human
CC vomeronasal organ (VNO) cDNA sequences. The cDNA library was
CC constructed from female tissue. Sequences isolated from the VNO cDNA
CC library are useful as a unique resource for the identification and
CC isolation of genes expressed in VNO, specifically genes for pheromone
CC receptors, ion channels and prospective reagents for high throughput
CC assays. They are useful as an excellent source to search for novel
CC genes, gene fragments or other nucleotide sequences encoding proteins
CC that are implicated in detection of pheromones or other vomeropherins
CC in the human VNO. The cDNAs are useful as hybridization probes for
CC determining the presence or concentration of an oligo- or polynucleotide,
CC e.g. DNA, of interest.
XX
SQ Sequence 628 BP; 138 A; 178 C; 123 G; 189 T; 0 other;

Query Match 14.8%; Score 164.8; DB 22; Length 628;
Best Local Similarity 60.3%; Pred. No. 3 6e-36;
Matches 381; Conservative 0; Mismatches 237; Indels 14; Gaps 6;

Qy 258 acatgataattagtcacctgtctctcatccacatactgtctctcactcagcgcaatat 317
Db 1 acctgatcatcagctctctgtggccctcatccaccttggatgc---taacagtcattggat 57
Qy 318 tgggtccttagactctcttgggttcacagaataactcaggtatgtcttaggtataaaggtca 377
Db 58 tcagagctgttgatatttttgcattctcagaatgtgtggaatgacatacaaatgcaatccc 117
Qy 378 ttgtctttttaacaagagtgatgagggcctctccactgtcacccctgctcctcctcagtg 437
Db 118 tgcaccattacacagacttttggagggcctctctcttgtgtacctgtctgtcagta 177
Qy 438 tgttcaggccatca---tcagcccacagatctctctcttgggcaaaagctcaaacatcctt 494
Db 178 tcttcaggccatccaccttagcccgagaagctcctgttagcaaaagtccaataataat 237
Qy 495 ctgcaagtcacatcttaggattctcttttctcatgggtcctcacaacatgttcatgtgtg 554
Db 238 ccacacagcacagcctgtgttccctctctgtctctgtggcctctctacatctcctgtgta 297
Qy 555 taatcttctgtgtac--actgcgggtaccccccagtgaaacggggccagctctctctgttg 612
Db 298 ctcaactctcttcacacctgtgtgtgactacaactctcttcacgcagctcatattg 357
Qy 613 tcatacagcactgtctcttttggccataggtacagctacacccacagagagactgttttcaac 672
Db 358 tcaatgaatctgcattattttaccatggattaccacccagggatttttttccatat 417
Qy 673 taatgactttgagggatgtcacctttatagggttcaggtctcctcacaagaggtcacatgg 732
Db 418 tggggatatttcgggatgtgtccttcatagggtctcagccctcctccagcgggtacatgg 477
Qy 733 tgattattttatcacagacaataaagaggtctatctcagtyccttccacacagccagctgtcc 792
Db 478 tggccctctgtgcagaca-caggaaacagggcccagcatcttcacagggatccagcctttct 536
Qy 793 cc--gagctcacagtgaaaagagcctccaggtccagctatcttactgtgtgag--tttt 847
Db 537 ccaaaagcatccccagagcaaggccaccagaccatcctgtgtgcatgagcttctt 596
Qy 848 gtcttcacatactgggtggactttacgttctc 879
Db 597 gtgtgagtactgtgtggactgcaccatc 628

RESULT 15
AAH74901
ID AAH74901 standard; DNA: 632 BP.
XX
AC AAH74901;
XX
DT 29-OCT-2001 (first entry)
XX
DE Probe used to isolate human VNO cDNA sequences.
XX
KW Human; vomeronasal organ; VNO; pheromone receptor; ion channel;
KW pheromone; vomeropherin; probe; ss.
XX
OS Homo sapiens.
XX
PN WO200161046-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-US05178.
XX
PR 17-FEB-2000; 2000US-0183128.
XX

PA (PHER-) PHERIN PHARM INC.

PI Herman RC, Berliner D:

WPI: 2001-522609/57.

Novel vomeronasal organ cDNA library useful for identification and isolation of genes expressed in vomeronasal organs, specifically genes for pheromone receptors and ion channels -

PS Example 9; Page 52-53; 59pp; English.

The present sequence represents a probe, used to isolate human vomeronasal organ (VNO) cDNA sequences. The cDNA library was constructed from female tissue. Sequences isolated from the VNO cDNA library are useful as a unique resource for the identification and isolation of genes expressed in VNO, specifically genes for pheromone receptors, ion channels and prospective reagents for high throughput assays. They are useful as an excellent source to search for novel genes, gene fragments or other nucleotide sequences encoding proteins that are implicated in detection of pheromones or other vomeropherins in the human VNO. The cDNAs are useful as hybridization probes for determining the presence or concentration of an oligo- or polynucleotide e.g. DNA, of interest.

Sequence 632 BP: 150 A: 177 C: 109 G: 196 T: 0 other:

Query Match	14.1%	Score 156.8;	DB 22;	Length 632;
Best Local Similarity	59.5%;	Pred. NO. 6.4e-34;		
Matches 357; Conservative	0;	Mismatches 232;	Indels 11;	Gaps 5;

Qy 261 tgataattagtcacctgtctctcatccacatactqctqctctctcactcagqcaatatattq 320

Db 1 tggccattgggtctcttattccctaatccacctactgatactactgataggggc---attca 57

Qy 321 tqtccttagactctcttttqqttcacagaaatactcagqatgattcttagqataaagqtcattg 380

Db 58 taqccataqacattttttatttcttqqagqqqatqqqatgacatcatatgtaaatttccttg 117

Qy 381 tcttttaaacaaqqtgatgagggqccctcccatctgcacccccctgcctcctgaqtatgc 440

Db 118 tctacttgtacagaaqttttaqaqqctctctcttcttaccacctqcatqttqaqatqccc 177

Qy 441 tccaqqccatca--tcaqcccccaqcatcttctccttqqcaaaqctcaaacatcccttctq 497

Db 178 **tg**caggccatcacccctcagccccagaaqctcctgttttaqcaaaqttcaaacataaqtctc 237

Qy 498 caaqtccacatcttaqqatcttccctttctcatqqgtccctcaacatgttctcatctggtgtaa 557

Db 238 cccatcacgtctcctgtgccattatttcgctgagcatcctctacatgtttcattagcaqtc 297

Qy 558 tcttctgctgtacact--gcqgctaccccccaqtgaaacqgqcccaqtcttctgtatttqtca 615

Db 298 acctcttaqtatcccatcaatqccaccccccaatttqaccacgaacaacttttatqcaagtta 357

Qy 616 tacaqcactgttcctttttgcccattgaqctacacccacaggaqactgtttttcacactaa 675

Db 358 ctcaqtccctacattatacccttgaqgttacctcatgcaaaagcatgttttctacacttc 417

676 **tgactttgagggaatgtcaccttttatagggttcattggtcctctcaaggaggtacatggtga** 735

Db 418 **tgccatcagagatatctctcttattagtctcatggtccctcgaactggttacatgagg** 477

0v 736 ttattttatacagacaaataaggagctatctcagtaccttcacacaccacacctatcccc- 794

Db 478 ttctcttqtataggc-acagggaatcagatccagcatcttcaaggagccaaacctttcccca 536

Ov 795 -gagtcaccagtgaaaaagagcctcccaggctatcttactactgagtgatttttgtcttc 853

Db 537 aaagcatctccagaacaaaggccacacagaccatcctgaatcctcatgaccttctttatc 596


```
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 138592
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-138592
```

```
Query Match 2.9%; Score 32; DB 6; Length 738;
Best Local Similarity 100.0%; Pred. No. 5.1e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 390 acaaggtgatgagggcctctccatctgcacc 421
|||||
Db 723 ACAAGGTGATGAGGGCCTCTCCATCTGCACC 692
```

```
RESULT 3
US-10-027-632-325100/c
; Sequence 325100, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 325100
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-325100
```

```
Query Match 2.9%; Score 32; DB 6; Length 738;
Best Local Similarity 100.0%; Pred. No. 5.1e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 390 acaaggtgatgagggcctctccatctgcacc 421
```

```
Db 723 ACAAGGTGATGAGGGCCTCTCCATCTGCACC 692
|||||
```

```
RESULT 4
US-10-027-632-324938
; Sequence 324938, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 324938
; LENGTH: 916
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-324938
```

```
Query Match 2.9%; Score 32; DB 6; Length 916;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 390 acaaggtgatgagggcctctccatctgcacc 421
|||||
Db 91 acaaggtgatgagggcctctccatctgcacc 122
```

```
RESULT 5
US-10-027-632-325047
; Sequence 325047, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 325047
```

LENGTH: 916
TYPE: DNA
ORGANISM: Human
US-10-027-632-325047

Query Match 2.9%; Score 32; DB 6; Length 916;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 acaaggtgatgagggcctctccatctgcacc 421
|||||
Db 91 acaaggtgatgagggcctctccatctgcacc 122

RESULT 6

PCT-US02-09923-131
Sequence 131, Application PC/TUS0209923

GENERAL INFORMATION:

APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: THORNTON, Michael
APPLICANT: YAO, Monique G.
APPLICANT: RICHARDSON, Thomas W.
APPLICANT: SWARNAKAR, Anita
APPLICANT: KALLICK, Deborah A.
APPLICANT: ISON, Craig H.
APPLICANT: WALIA, Narinder K.
APPLICANT: GANDHI, Ameena R.
APPLICANT: LEE, Ernestine A.
APPLICANT: ELLIOTT, Vicki S.
APPLICANT: HAFALIA, April J.A.
APPLICANT: AU-YOUNG, Janice
APPLICANT: GRIFFIN, Jennifer A.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: KHAN, Farrah A.
APPLICANT: BECHA, Shanya
APPLICANT: LU, Yan
APPLICANT: ARVIZU, Chandra P.
APPLICANT: BOROWSKY, Mark L.
APPLICANT: LAL, Preeti G.
APPLICANT: RAMKUMAR, Jayalaxmi
APPLICANT: EMERLING, Brooke M.
APPLICANT: WALSH, Roderick T.
APPLICANT: YUE, Henry
APPLICANT: BURFORD, Neil
APPLICANT: GRUL, Richard C.

TITLE OF INVENTION: G-PROTEIN COUPLED RECEPTORS

FILE REFERENCE: PF-0933 PCT

CURRENT APPLICATION NUMBER: PCT/US02/09923

CURRENT FILING DATE: 2002-03-29

PRIOR APPLICATION NUMBER: 60/280,683; 60/283,714; 60/287,266; 60/285,336

PRIOR FILING DATE: 2001-03-30; 2001-04-13; 2001-04-27; 2001-04-20

NUMBER OF SEQ ID NOS: 146

SOFTWARE: PERL Program

SEQ ID NO 131

LENGTH: 1097

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Incyte ID No: 7476246CBI

PCT-US02-09923-131

Query Match 2.9%; Score 32; DB 1; Length 1097;
Best Local Similarity 100.0%; Pred. No. 5.3e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 acaaggtgatgagggcctctccatctgcacc 421
|||||
Db 351 acaaggtgatgagggcctctccatctgcacc 382

RESULT 7

PCT-US02-09188-1323
Sequence 1323, Application PC/TUS0209188
GENERAL INFORMATION:

APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Human Secreted Proteins
FILE REFERENCE: PS952PCT
CURRENT APPLICATION NUMBER: PCT/US02/09188
CURRENT FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: US 60/278,650
PRIOR FILING DATE: 2001-03-27
PRIOR APPLICATION NUMBER: US 09/950,082
PRIOR FILING DATE: 2001-09-12
PRIOR APPLICATION NUMBER: US 09/950,083
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 1732
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1323

LENGTH: 83

TYPE: DNA

ORGANISM: Homo sapiens

PCT-US02-09188-1323

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 392 aaggtgatgagggcctctccatctgcacc 421
|||||
Db 1 aaggtgatgagggcctctccatctgcacc 30

RESULT 8

PCT-US02-09188-1326
Sequence 1326, Application PC/TUS0209188
GENERAL INFORMATION:

APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Human Secreted Proteins
FILE REFERENCE: PS952PCT
CURRENT APPLICATION NUMBER: PCT/US02/09188
CURRENT FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: US 60/278,650
PRIOR FILING DATE: 2001-03-27
PRIOR APPLICATION NUMBER: US 09/950,082
PRIOR FILING DATE: 2001-09-12
PRIOR APPLICATION NUMBER: US 09/950,083
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 1732
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1326

LENGTH: 83

TYPE: DNA

ORGANISM: Homo sapiens

PCT-US02-09188-1326

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 392 aaggtgatgagggcctctccatctgcacc 421
|||||
Db 1 aaggtgatgagggcctctccatctgcacc 30

RESULT 9

PCT-US02-09188-1342
Sequence 1342, Application PC/TUS0209188
GENERAL INFORMATION:

APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Human Secreted Proteins
FILE REFERENCE: PS952PCT

; CURRENT APPLICATION NUMBER: PCT/US02/09188
; PRIOR FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1732
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1342
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09188-1342

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 aagtgatgagggcctctccatctgcacc 421
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 10
PCT-US02-09188-1343
; Sequence 1343, Application PC/TUS0209188
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS952PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09188
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1732
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1343
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09188-1343

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 aagtgatgagggcctctccatctgcacc 421
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 11
PCT-US02-09370-1381
; Sequence 1381, Application PC/TUS0209370
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS954PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09370
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083

; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1834
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1381
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09370-1381

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 aagtgatgagggcctctccatctgcacc 421
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 12
PCT-US02-09370-1384
; Sequence 1384, Application PC/TUS0209370
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS954PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09370
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1834
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1384
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09370-1384

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 aagtgatgagggcctctccatctgcacc 421
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 13
PCT-US02-09370-1400
; Sequence 1400, Application PC/TUS0209370
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS954PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09370
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1834
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1400
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens

PCT-US02-09370-1400

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 392 aagtgatgagggcctctccatctgcacc 421
|
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 14

PCT-US02-09370-1401
; Sequence 1401, Application PC/TUS0209370
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS954PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09370
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1834
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1401
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09370-1401

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 392 aagtgatgagggcctctccatctgcacc 421
|
Db 1 aagtgatgagggcctctccatctgcacc 30

RESULT 15

PCT-US02-09922-872
; Sequence 872, Application PC/TUS0209922
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS955PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09922
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 1117
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 872
; LENGTH: 83
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-09922-872

Query Match 2.7%; Score 30; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 392 aagtgatgagggcctctccatctgcacc 421
|
Db 1 aagtgatgagggcctctccatctgcacc 30

Search completed: July 21, 2002, 05:31:07
Job time: 5576 sec

